Proceedings of the 23rd Amsterdam Colloquium



23rd Amsterdam Colloquium *2022*

Edited by Marco Degano, Tom Roberts, Giorgio Sbardolini &
Marieke Schouwstra

Foreword

The papers contained in this collection were presented at the 23rd Amsterdam Colloquium, organised by the Institute for Logic, Language and Computation (ILLC) at the University of Amsterdam, December 19–21, 2022. The Amsterdam Colloquia aim at bringing together linguists, philosophers, logicians, cognitive scientists and computer scientists who share an interest in the formal study of the semantics and pragmatics of natural and formal languages.

Besides the general programme, the 2022 edition featured two workshops on *Biases in Language and Cognition* and on *Compositional Approaches to Projection*, and the *Beth Lecture*, jointly organised with the E.W. Beth Foundation. The programme included 8 invited talks, 34 contributed talks, and a poster session with 25 contributed posters.

We would like to thank all programme committee members, listed below, and all reviewers for their efforts in selecting the contributed talks and posters. We would also like to thank Peter van Ormondt, Ewout Arends and Katrin Schulz for their practical help in organising the conference.

Lastly, we would like to thank the ILLC, the E.W. Beth Foundation, the Netherlands Organization for Scientific Research (NWO), and the European Research Council (ERC, grant no. 819455) for financial support.

Marco Degano
Tom Roberts
Giorgio Sbardolini
Marieke Schouwstra

Programme Committees

General programme

Maria Aloni	ILLC, University of Amsterdam
Paul Dekker	ILLC, University of Amsterdam
Luca Incurvati	ILLC, University of Amsterdam
Floris Roelofsen (chair)	ILLC, University of Amsterdam

Workshop on Biases in Language and Cognition

Marieke Schouwstra	ILLC, University of Amsterdam
Katrin Schulz	ILLC, University of Amsterdam

Workshop on Compositional Approaches to Projection

Reinhard Muskens	ILLC, University of Amsterdam
Tom Roberts	ILLC, University of Amsterdam
Émile Enguehard	ILLC, University of Amsterdam

Table of Contents

Invited

Raising hell: On the hyper-projectivity of slurs 1
David Beaver and Jason Stanley

General Programme

Not all scalar inferences are alike: The effect of existential presuppositions 8
Omri Amiraz

Disentangling semantic empty-set effects in quantifier comprehension from simple associations: processing evidence from exceptive-additives 15
Nadine Balbach, Fabian Schlotterbeck, Hening Wang and Oliver Bott

Positivity, (anti-)exhaustivity and stability 23
Moshe E. Bar-Lev and Roni Katzir

Expressives do shift, but not on their own 31
Corien Bary, Leopold Hess and Bob van Tiel

As complex as they appear: Children's comprehension of conjunctive expressions in Georgian 38
Cory Bill, Aurore Gonzalez, Imke Driemel and Tamar Makharoblidze

Antipresuppositions, logophors and shifted indexicality 45
David Blunier

The semantics of Rudin constructions in Romanian 55
Ivano Caponigro and Anamaria Falaus

Future-past asymmetries, evidential grounding, and projection 62
Fabrizio Cariani

Disambiguating will-conditionals 69
Ivano Ciardelli and Adrian Ommundsen

A minimally sufficient analysis of sufficiency modal constructions 77
Cleo Condoravdi and Itamar Francez

A scalar analysis of polarity contrasts between exclusive modifiers 84
Lucas Fagen

Predicate Connection and Informational Oddness 91
Diego Feinmann

Clausal embedding under to in Japanese as speech acts 98
Daniel Goodhue and Junko Shimoyama

Bias in high negation questions as a quantity implicature in commitment space semantics 106
Daniel Goodhue

When maximality is not maximally informative: A compositional analysis of the depth-charge illusion 113
Vera Hohaus and Nadine Bade

Compatibility, compossibility, and epistemic modality . . . 120
Wesley Holliday and Matthew Mandelkern

The Hell with questions . . . 127
Michela Ippolito

Neg-raising requires stativity: A scaleless implicature based explanation . . . 134
Paloma Jeretič and Deniz Özyıldız

Syntactic freezing as semantic matching: licensing syntactic deviation in echo questions . . . 145
Jing Ji

Plural priming revisited: inverse preference and spillover effects . . . 152
Yizhen Jiang, Shiyu Ren, Yihang Shen, Richard Breheny, Paul Marty and Yasutada Sudo

The Acquaintance Inference as Scalar Exhaustification. . . . 159
Jonathan Kendrick

Investigating the properties of clause-embedding predicates in Polish . . . 166
Tomasz Klochowicz

Modeling Metadisputes . . . 173
Alexander Kocurek

The Role of Questions, Circumstances, and Algorithms in Belief . . . 181
Alexander Kocurek, Jens Kipper and Zeynep Soysal

Alternate Categorization: a Formal-Conceptual Semantics of Reciprocal Alternations . . . 188
Imke Kruitwagen, James Hampton, Yoad Winter and Joost Zwarts

Aboutness and Modality . . . 194
Dean McHugh

An existential semantics for habituals . . . 207
Jéssica Mendes

Two Types of Intensional Transitive Verbs: The view from Farsi . . . 214
Esmail Moghiseh and Luis Alonso-Ovalle

Dynamic Relationism about Belief . . . 221
Dilip Ninan

Non-past directives in Japanese . . . 228
Yuya Noguchi

Subatomic and plural homogeneity as exhaustification effects of different kinds . . . 236
Mathieu Paille

Numerous relative clauses: permutation invariance, anti-restrictiveness, triviality . . . 244
Lorenzo Pinton

Diversity with Universality . . . 251
Sonia Ramotowska, Paul Marty, Jacopo Romoli, Yasutada Sudo and Richard Breheny

Exclusives vary in strength and scale structure: experimental evidence . . . 258
Eszter Ronai and Lucas Fagen

Ability Modals as Causal Modals 267
Paolo Santorio

Breaking the Fourth Wall and (Meta)fictional Reference 275
Merel Semeijn

Projective variability of (semi-)factive verbs in family of sentence contexts: A rating study ... 282
Judith Sieker and Torgrim Solstad

Bare nominals are underspecified for (in)definite meanings, not ambiguous between them 290
Sadhwi Srinivas

Contrafactives and learnability 298
David Strohmaier and Simon Wimmer

Morphological effects on indexical shift in Uyghur 306
Ruoan Wang

Revisiting presuppositional accounts of homogeneity 313
Jad Wehbe

Mandarin-speaking children's understanding of you 'again' with goal-PPs 320
Ting Xu, William Snyder and Stella Christie

Epistemic Pseudo-Agreement and Hyper-Domain Semantics 327
Yichi Zhang and Eric Pacuit

*Be surprised whether: question-to-cleft reduction 336
Ziling Zhu and Yimei Xiang

Workshop: Biases in Language and Cognition

Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity 344
Milica Denić and Jakub Szymanik

Genericity and the too-many-structures puzzle: Insights from an experimental study 353
Imke Driemel, Johannes Hein, Desiré Carioti, Jakob Wünsch, Vina Tsakali, Artemis Alexiadou, Uli Sauerland and Maria Teresa Guasti

In search of a unified mechanism for regularisation of linguistic variation 360
Aislinn Keogh, Simon Kirby and Jennifer Culbertson

New evidence for the unlearnability of non-conservative quantifiers 367
Tyler Knowlton, John Trueswell and Anna Papafragou

Generalizing polarity sensitive items to positive environments in an Artificial Language 375
Jeremy Kuhn and Mora Maldonado

Testing communicative and learning biases in a causal model of language evolution: A study of cues to Subject and Object 383
Natalia Levshina

Interpretations of plurality in personal pronouns differ across person categories: An experimental study 388
Mora Maldonado and Carmen Saldana

Workshop: Compositional Approaches to Projection

Cataphoric presupposition verification and symmetric filtering: The case of occasion verbs ... 397
Oliver Bott and Torgrim Solstad

A constraint on presupposition accommodation 405
Omri Doron and Jad Wehbe

An alternative semantics for presupposition 412
Julian Grove

Presupposition projection from the scope of quantifiers: A novel approach 422
Alexandros Kalomoiros

Preconditions, Presuppositions, and Projection 429
Craige Roberts and Mandy Simons

Clause-internal coherence as presupposition resolution 437
Kelsey Sasaki and Daniel Altshuler

Presupposition Triggering Reflects Pragmatic Reasoning About Utterance Utility 444
Alex Warstadt

Raising hell:
On the hyper-projectivity of slurs

David Beaver[1] and Jason Stanley[2]

[1] The University of Texas at Austin
dib@utexas.edu
[2] Yale University
jason.stanley@yale.edu

Abstract

Mere mentions of a slur have the potential to do harm, assaulting a listener, perpetuating marginalization of oppressed groups, and solidifying oppressive practices. Whereas most projective meanings, such as presuppositions, are typically assumed to be blocked by quotation, some effects of slurs are found even when the slur is quoted. In this sense, slurs are *hyper-projective*. Our goal is to explain why. The material here is drawn from: David Beaver and Jason Stanley, *The Politics of Language*, Princeton University Press (in press), primarily from Chapter 10, "Oppressive Speech."

The resonances of swearwords and slurs (particularly when used in acts of slurring) inflict themselves painfully on hearers, and, like an inebriated gatecrasher at a party, draw collective attention, brash, enervating, and obdurate. These effects seem almost magical. How do we explain this in terms of the conventional properties of those words? We do not think slurs are magical. Quite the opposite. There is, unfortunately, no language more ordinary than slurs.

At the very least, a theory of slurs must account for three properties. Slurs are:

Ideologically revelatory: bringing a "complex of historical facts and constructed attitudes [...] to attention" although it's "hard [...] to discern exactly what those attitudes are, or what the precise historical facts being deployed are", as McCready and Davis [7] put it.

Exigent: forcibly impacting hearers' emotional states.

Hyper-projective: blurring the distinction between use and meaning.

We focus in this extended abstract on the last of the three properties.

Let us first briefly state the features of the framework we introduce in *The Politics of Language* [3], in terms of which our account of slurs will be stated. Words have *resonances*, all the things that are found in the contexts in which they are used. The degree to which a feature of context is a resonance of a communicative behavior is proportional to the boost in probability of the feature given that behavior. People and collections of people have *attunements*, which are ways of behaving, thinking and feeling about things. When people have attunements to a word, that means not only that they have theoretical knowledge of what the word means and how it fits into the grammar, but also that they have practical knowledge. That means that both their use of the word and their reactions to use of the word are appropriate for the community of practice within which the word is found. As we have set things up, resonance and attunement are two sides of the same coin. To be attuned to any practice is to have a tendency for change of behavior and state in accord with the resonances of the practice. Our notion of *harmony* concerns the fact that people feel dissonance when they sense a clash of salient attunements, and consonance when they feel alignment and coherence. As we have defined it, an *ideology* is a system of collective attunements of a community of practice, a system which the community has harmonized around. An important sub-type of ideologies for current

purposes are *discriminatory ideologies*, those that include attunements to in-group/out-group distinctions, and in which members of out-groups are valued less than members of in-groups, and hence as inherently deserving of less than equal treatment or resources.

Let us turn now to *hyper-projectivity*, which, as the name suggests, is a special type of *projectivity*. Projectivity is commonly thought of in terms of special inferences associated with expressions. What makes the inferences special is that they are found even when the expressions they are associated with are embedded under operators that tend to block ordinary entailments. For example, conditionals block entailments, so it is no surprise that the entailment from "Your car is worth over $5000" to "Your car is worth over $4000" is blocked when inside a conditional; the latter does not follow from "If your car is worth over $5000, you should sell it." However, another inference is associated with "Your car is worth over $5000", namely that the addressee has a car. This inference is not blocked by the conditional, since it follows from "If your car is worth over $5000, you should sell it." that the addressee has a car. Accordingly, the hearer's ownership of a car can be described as a projective inference. In this case, the explanation is standardly said to be that the possessive "your car" is a presupposition trigger, and the inference that the addressee has a car is said to be a presupposition.

The details of how presuppositions project in meta-linguistic environments (for example from quotative contexts) remain under-explored. Perhaps that is because projection in such environments has been assumed to be so limited that it cannot usefully be used to diagnose presupposition. What is empirically clear is that while presupposition projection is standardly taken to be more-or-less blocked by meta-linguistic embedding, some resonances of slurs project strongly from meta-linguistic environments. This is in fact the reason we, like many authors, avoid even mentioning the N-word (as opposed to referring to it indirectly with the awkward locution "the N-word"). Hyper-projective resonances are those which are clearly not plugged, but project from meta-linguistic environments.

Note that we speak of resonances rather than inferences. The idea is that neither the gut-punch effect on a target, nor the social-positioning effect, are simply propositions that a speaker intends to convey. Furthermore, we do not think it would be appropriate to term these effects *perlocutionary*. That would suggest that the function of the slur was to convey a certain proposition, and then there just happened to be some pragmatics that a sly speaker could exploit in order to strategically engineer negative effects on those slurred. While we do not argue the case in this short paper, we take it that both gut-punch and social-positioning effects are central to a slur's conventionalized function. Although our notion of what it is to be part of an expression's conventionalized meaning is itself somewhat non-standard, it will suffice for current purposes to think of the resonances of an expression as all aspects of the context that are to some extent conventionally present when the expression is used. These aspects of context include those that are effects of the use of the expression.

Some theories of slurs explain their effects in terms of presuppositions, and still others explain those effects as conventional implicatures. As regards hyper-projectivity, both presuppositional accounts and conventional implicature accounts make some headway. But it is hard for extant presuppositional accounts to explain why some presuppositions should be more projective than others. And while it is well established that conventional implicatures tend to be more projective than most presuppositions, neither conventional implicature accounts nor prior presuppositional accounts can explain why mere mentions should lead to any projection at all.

Now let us turn to our own theory of slurring. Here are the main points:

1. Slurring labels a target using a negatively evaluative term, a slur.
2. The label identifies the target with a disdained, despised or hated out-group category associated with stereotypical properties within a discriminatory ideology. Stereotypical

properties might include submissiveness, voicelessness, or technical incompetence.

3. The discriminatory ideology is presupposed by the slurring action. Labeling someone using the slur category draws attention to those aspects of the discriminatory ideology which define the category, and to any differences between the slurring ideology and other salient ideologies (such as that of a co-existing non-discriminatory ideology).
4. The resonances of the slur include these attentional and emotional effects on the target group, a range of attunements belonging to the presupposed discriminatory ideology, a demarcation of in-group and out-group, and power differentials between these groups.
5. For an in-group member, attunement to the slurring practice includes feeling such emotions as hatred, superiority, and alignment with the in-group when involved in a slurring interaction as speaker or hearer.
6. For an out-group member exposed to the slurring practice, attunement implies experience of painful loss of face, painful confrontation with the discriminatory ideology to which attention has been drawn, and further dissonance. The dissonance involves a dramatic conflict between desired private face and public face, and in many cases between a desired way of life and life as a member of an oppressed group.

As an example, let's pick the relatively mild slur "muppet", which is conventionalized in the UK to suggest something similar to "airhead" in the US. The resonances of "muppet" build on associations with the soft puppets that came to fame in the US children's show Sesame Street, and are suggestive of someone who tends not to act or speak in an incisive, intelligent, or self-directed way. The resonances of "muppet" include not just the features that an individual so-described might be expected to have, but also emotions, dispositions, and cognitive attitudes.

The emotions, dispositions, and cognitive attitudes on the speaker's side are different than those on the target's side. Emotionally, it is clear that the speaker looks down on the intelligence of the target. The speaker is expected to feel superior, and the target inferior. Dispositionally, the resonances include ways of using "muppet", and ways of reacting to it. A resonance of the term is that it is not simply used to insult people, although it can be, but is also used to upbraid people in a mildly affectionate manner - "You muppet!" To be attuned to the UK "muppet" practice is to use it appropriately, and to react to its use appropriately, for example not feeling cut to the bone when somebody calls you a muppet, but rather feeling a little sheepish about whatever unfortunate thing you have done without thinking properly.

The ideology associated with slurring use of "muppet" includes various practices of using the word, but also ideas about the type of intelligent and independent behavior that can be expected of people. Insulting uses of the term betray people's values: the slurring use of "muppet" always belittles the target's competence, and a harsher use ("Get the f- out of my way you stupid f-ing muppet"), although far from being a strong slur, further betrays a perceived differential in competence, and possibly power, between speaker and target.

Similar remarks apply to the Southern US use of "precious", in phrases like "Aren't you precious!", except that here being fully attuned to it means reacting to its use in a way that reflects the fact that you are probably not being complimented, but insulted. Southern US "precious" and other expressions which damn with praise ("Bless your heart!") are associated with an ideology in which a distinction is made between two types of people: those that have got it together and keep things running as they ought, and a second group about whom the best we can say is that they are, e.g., cute, or at least deserving of the good lord's love, but are functionally less than fully competent. It is important to our analysis that slurs do not merely label, and do not merely draw attention to ideology, but simultaneously produce negative and often painful emotions. Slurs destroy self-image, publicly shame, and can thereby also silence.

They have power. To resist the immediate attentional and emotional effects of the slur is difficult. But it is certainly possible to actively resist the silencing effect of slurs. In reaction to the use of the N-word by white TV personality Bill Maher, the rapper Ice Cube, who is Black, compared the power of the N-word to the stab of the knife, which wounds independently of the intentions of the speaker: "You know, it's a word that has been used against us. It's like a knife, man. You can use it as a weapon or you can use it as a tool. [....] It's not cool because when I hear my homie say it, it don't feel like venom. When I hear a white person say it, it feel like that knife stabbing you, even if they don't mean to." (Real Time with Bill Maher", 9 June 2017) [5].

The primary point of names is to draw attention to individuals. Descriptive names like vocative uses of the appellations "Professor" or "Private," simultaneously draw attention both to a person and to a role. The attention any name draws to an individual is hyper-projective. Even if one uses a name within quotation marks, the audience's attention is still drawn to its referent. In a similar way, the function of insults is to draw attention to despised characteristics of an individual. To slur is to label an individual or group with a despised role within a discriminatory ideology. Note that simply naming an ideology draws attention to the ideology, but is not offensive. We can refer to fascism, racism, or sexism without offending, unless someone thinks that by doing so we are also labeling them as having a despised ideology. Using the terms does not necessarily cause anyone to lose public face. The victim of a name-calling act of slurring loses public face, and will recognize that public loss of face to the extent that they are attuned to the utterance situation. The offensiveness and hurtfulness of a slur depends on its drawing attention to aspects of an ideology which are offensive to a person who is slurred, because the ideology debases that person, and normalizes oppressive practices against them.

We take the hyper-projectivity of slurs to consist in the fact that mere mentions of some slurs, including quotative uses, can cause offense. The N-word is the clearest case. Here is Jyoti Rao, an academic psychotherapist, discussing situations in which the N-word is mentioned or quoted as an object of study at academic conferences [8]:

> What happens after the slur is uttered has been equally consistent, and equally notable. Some small part of the group, typically comprising the few people of color and others present who come from marginalized backgrounds, attempt to bring attention to what they have just experienced. Often, they express palpable pain, clearly expressing the destabilizing effect of hearing a word strongly associated with white supremacy spoken at a conference. Aside from occupying a marginalized social identity, those speaking are frequently earlier in their career; hold positions as graduate students, junior faculty, or analytic candidates; are less financially secure; and are speaking to people with greater institutional, organizational, and other forms of power. In response to hearing from these participants, the user of the word, and several other white people in the group, seem mobilized to counter what has been reported about the consequences of the epithet's use. They begin to explain that speaking the n-word is benign, even salutary, and advocate for why the word should be used freely by white people and psychoanalysts. In all the cases I have seen, the person using the word, as well as the people roused in support, have appeared unmoved, unreceptive, and unapologetic, even when it is repeatedly pointed out to them that their speech has caused harm.

Mere mentions of a slur have the potential to do harm. They perpetuate marginalization of oppressed groups, and solidify oppressive practices.

Note here that the philosophy and linguistics literature is peppered with slurs. We take it

that those quoting or mentioning slurs in this literature are not being intentionally vindictive. Rather, it is the result of an error, due to the acceptance of a certain interpretation of the use-mention distinction, a distinction at the heart of contemporary analytic study of meaning. According to this dogma, merely mentioned or quoted language is in some sense inert. It is as if quotes or italics were a lead-lined box, so the scholar who uses these devices cannot possibly be risking harm with the radioactive material inside. But theoretical prejudice should not interfere with reality. So let us just say this. How can the scholar know that their belief that a quoted slur will cause no harm is correct? If some evidence should arise suggesting that someone was harmed by exposure to a mere mention, would that not indicate that maybe it would be better to treat at least certain slurs a little more carefully, perhaps being very sure that there is no reasonable alternative to mentioning them before doing so?

Even if, like Bianca Cepollaro [4], one has a theory that treats slurs as not capable of being contextually cancelled, it remains mysterious, on any standard account, why even quoting slurs should be objectionable, or how it could possibly be a source of harm. We can explain this straightforwardly in terms of attention. Drawing attention to something does not require use. The mere mention of a word can suffice to draw attention to something. If in a report on an important news event, say concerning revelations of spying, one paragraph includes a quote from an anonymous official that mentions you by name, you are sure to soon find reporters at your door. By mentioning you, the report has drawn attention to you. If someone mentions the combination of a safe in a quote, e.g. "I heard Marie muttering the number 1673", they have drawn attention to the number, which may or may not have security implications.

Let us briefly head off a potential objection to our suggestion that attention is primary. Some might point out that when something is mentioned, the Gricean maxim of Relevance can then explain the communicative significance of the mention. On this view, when a hearer encounters a mention of a term, they must then reconstruct the intention of the speaker in mentioning whatever it is. The reason why Jason keeps mentioning figure skating news, the hearer might reason, is because he wants people to think about his recent Olympic figure skating success. We completely agree that such reasoning processes occur, and that they can explain aspects of the communicative significance of mentions, including mentions of slurs.

Joe Biden once caricatured former New York City Mayor Rudolph Guiliani by saying "There's only three things he mentions in a sentence – a noun, a verb, and 9/11." What he was referencing was Giuliani's habit of drawing attention to the terrorist destruction of the World Trade Center, and hence his role in the city's recovery. Although it might be argued that Giuliani was simply obsessed by the event, it seems more plausible that, as Biden was implicitly suggesting, Giuliani was strategically mentioning 9/11. Donald Trump has a habit of beginning sentences with "People say...." and related locutions. The discursive logic of this strategy is multifaceted, but at least part of it is that allows Trump to draw attention to ideas without taking responsibility for them. An authoritarian leader needs to be a master of attention. Sometimes the strategic point of an utterance is to draw attention to something.

However, as an explanation of the direct emotive and attentional effects of slurs, such a strategic explanation is backwards. Relevance does not explain attentional effects. It's rather the case that a premise of relevance-based argumentation is that the speaker has drawn attention to something. The question the Gricean theorist asks is: why has the speaker drawn attention to it? The question we ask is: given that the speaker has drawn attention to something, what effects will that have independently of what the speaker's intention was. We do not dispute that there are effects that should be explained in terms of the speaker's intention, or indeed that such reasoning is important in considering what form of counterspeech is appropriate (e.g., blame and censure vs. education). What we dispute is that considerations of the speaker's

motivations are needed to explain the exigent power of the slur, its hyper-projectivity, or its ideological revelatoriness. And our explanation for hyper-projectivity centers on a claim that we take to be somewhat self-evident: **the attention drawing power of any construction is not plugged by quotation, or, for that matter, by other meta-linguistic embeddings**.

Uses of slurs in labeling and address practices are associated with a diverse mixture of other practices that form part of the oppressive ideology to which the slur belongs. For example, the use of a racial slur might be associated with practices of commanding, prohibiting, punishing, or blaming. All of these further practices have linguistic dimensions, but their existence does not need to be explained by a special theory of the linguistic properties of slurs. It is rather the case that part of the power of (some) slurs is derived through the resonances they have due to other practices that they are associated with. Using a slur to name-call can presuppose the right of the speaker, and all within the speaker's in-group, to command, prohibit, punish or blame the addressee, and all within the addressee's in-group. These are powerful presuppositions. But again, one does not need a special theory of slurs, or of presupposition, to explain that power: the power is derived from the system of oppression within which the slur is set, combined with the far more general account of social attunement and presupposition.

In a series of papers, Luvell Anderson, Ernie Lepore and Matthew Stone [1, 2, 6] explain various properties of slurs, including hyper-projectivity, as resulting from the existence of taboo. Leaving aside the question of whether this counts as an explanation (one might worry that the "explanation" is that it is offensive to mention slurs because they are unmentionable without causing offense), or whether it can possibly explain which specific resonances project and which don't, there is something right about it. The fact that something is taboo means that mentioning it will immediately catch listeners' attention, and indeed lead to the suspicion that the purpose of mentioning it was to call it to the listeners' attention. Perversely, the social development of taboos around slurs strengthens their effectiveness, by increasing their attention-drawing power. However, the development of these taboos does not so much explain anything about slurs, as add to the complexity of generalizing. For a taboo is itself an idiosyncratic complex of practices. These practices include counterspeech, censorship and education, all of which are applied in different ways for different slurs, and for which their application varies over time. What have been deemed as acceptable mentions of the N-word have evolved continuously during our own lifetimes. We do not have hard data on the nature of any changes in which uses or mentions of the N-word by which people offend which other people in what way and to what extent, but we feel confident in saying that the facts of the matter have not been static.

The non-taboo nature of "muppet" and the oft-slurring adjective "precious" is illustrative of a problem with using taboo as the basis of an explanation, rather than something that itself needs to be theorized, because the taboos relevant to even the paradigmatic cases of racial slurs are complex. It would be at the very least a gross oversimplification to maintain that in the reconstruction-era South the N-word was taboo, when an officer of the law could openly use it as a term of address. It was a word in general use in a powerful community of practice. And it was a powerful word. How can the power of a word, it's *exigence* as we have termed it, be explained by the illicitness of using it if those with power feel its use is licit? Similarly, the *revelatoriness* of slurs, their ability to bring to the fore an ugly discriminatory ideology, is not explained by their taboo nature, but is rather largely responsible for the imposition of societal taboos. Last, the *hyper-projectivity* of slurs is not explained by their taboo nature. Rather, hyper-projectivity is part of what makes the taboos against using some slurs so sweeping.

References

[1] Luvell Anderson and Ernie Lepore. Slurring words. *Noûs*, 47(1):25–48, 2013.

[2] Luvell Anderson and Ernie Lepore. What did you call me? slurs as prohibited words setting things up. *Analytic Philosophy*, 54(3):350–63, 2013.

[3] David Beaver and Jason Stanley. *The Politics of Language*. Princeton University Press, In press. Projected publication: Fall 2023.

[4] Bianca Cepollaro. In defence of a presuppositional account of slurs. *Language Sciences*, 52:36–45, 2015.

[5] Robert Henderson, Peter Klecha, and Elin McCready. Response to Pullum on slurs. *LanguageLog*, July 20, 2017.

[6] Ernie Lepore and Matthew Stone. Pejorative tone. *Bad words: Philosophical perspectives on slurs*, pages 134–53, 2018.

[7] Elin McCready and Christopher Davis. An invocational theory of slurs. *Proceedings of LENLS*, 14, 2017.

[8] Jyoti M Rao. Observations on use of the n-word in psychoanalytic conferences. *Journal of the American Psychoanalytic Association*, 69(2):315–341, 2021.

Not all scalar inferences are alike: The effect of existential presuppositions*

Omri Amiraz

The Hebrew University of Jerusalem, Jerusalem, Israel
omri.amiraz@mail.huji.ac.il

Abstract

This paper investigates a novel puzzle regarding the indirect scalar inference (SI) 'not all' ⇝ 'some'. When 'not all' is embedded in a Strawson-Downward-Entailing context, some constructions give rise to more robust SIs than others even though they express the same proposition with the same lexical items. I argue that this contrast is related to whether or not the quantifier phrase occurs in a conventionalized focus construction. If it does, the SI is not derived by implicature calculation but arises from an existential presupposition triggered by the focus construction. Hence, it is not sensitive to contextual considerations.

1 Introduction

In this paper, I observe a puzzling pattern in the context of the indirect scalar inference (SI) 'not all' ⇝ 'some'. Consider the contrast between (1a) and (1b), where 'not all' is embedded under the factive Strawson-Downward-Entailing (SDE) predicate *surprised* [19]. (1b) strongly implies that some of the students failed even though this SI conflicts with the common ground (CG) that none failed. (1a), by contrast, does not have a 'not all but some' interpretation in this context, which makes it felicitous.

(1) *Context: Mary teaches Intro to Statistics. She accidentally gave the students the exam of Advanced Statistics. She expected they would all fail, but surprisingly, they all passed.*
Mary: You should all be proud of yourselves for passing the exam. Frankly, ...
a. I was surprised that you didn't all fail!
b. #I was surprised that not all of you failed!

Another example is in (2), where 'not all' occurs in an antecedent of a counterfactual conditional, which is a non-factive SDE context [19]. In both sentences, we consider counterfactual situations in which it is not the case that everyone died in the plane crash. However, only in (2a) are there also relevant situations where in fact nobody died, possibly because the plane never crashed (i.e., there is no 'not all but some' SI in (2a)). In contrast, (2b) only considers situations where some team members did die in the crash, which is incongruent with the consequent (because it is unlikely that the survival of some (non-specific) team members would suffice to guarantee winning the title). Thus, (2b) has a surprisingly robust SI, just like (1b).

(2) a. This team would have won the title if they hadn't all died in that plane crash.
b. #This team would have won the title if not all of them had died in that plane crash.

Note that the absence of an SI in (1a) and (2a) is what one would expect based on the behavior of other scalar expressions in this environment. For instance, 'or' sometimes triggers an SI (i.e.,

*This paper has grown out of joint work with Eunsun Jou, and I am grateful to Eunsun for her contribution. I also thank Nora Boneh, Omri Doron, and Todd Snider for their useful comments and suggestions.

'A or B' $\rightsquigarrow$ 'not both A and B') in factive SDE contexts (aka "negative factives") such as the complement of *surprised* or *unaware* [9, 16], e.g., (3). However, this SI is not generated if it conflicts with the CG, e.g., (4). Hence, the goal of this paper is to explain why sentences like (1b) and (2b) trigger an SI that is not sensitive to contextual considerations.

(3) John is unaware that his son used his credit card or payment app to buy video games.
$\rightsquigarrow$ John's son used his father's credit card or payment app, but not both.

(4) Nintendo is completely unaware that I or my stories exist.
$\not\rightsquigarrow$ I or my stories exist, but not both.

I argue that the contrast in (1) and (2) is related to whether or not the universal quantifier phrase (QP) occurs in a conventionalized negative focus construction (NFC). When 'all' occurs under negation in a construction which is not an NFC—e.g., in (1a) and (2a)—the 'not all but some' SI is derived by implicature calculation. Hence, it patterns with the SI of 'or' (3)-(4) and other scalar expressions (see Section 4 for additional examples). However, when 'all' occurs under negation in an NFC—e.g., in (1b) and (2b)—the SI is not derived by implicature calculation but rather arises from an existential presupposition triggered by the NFC (i.e., NFCs are presupposition triggers just like clefts). Given that the SI arises from a presupposition, it is present even when it contradicts the CG (1b) and in non-factive SDE contexts (2b).

This analysis suggests that some indirect SIs constitute a phenomenon in their own right. Additionally, previous studies have shown that some scalar items trigger SIs more frequently than others [17]. This paper observes that there is variation not only across lexical scales (e.g., ⟨some, all⟩ vs. ⟨warm, hot⟩, see e.g., [17]) but also between grammatical constructions expressing the same proposition and containing the same set of lexical items, as in (1) and (2).

This paper is structured as follows. Section 2 discusses the distinction between clausal negation constructions and NFCs and provides cross-linguistic evidence for its relevance to SIs. Section 3 proposes an information-theoretic analysis. Section 4 briefly discusses the behavior of other scalar expressions in factive SDE contexts. In contrast to previous claims in the literature, in particular in [16], I demonstrate that SIs are merely optional in this environment and argue that this fact poses a challenge for certain previous accounts of SIs.

2 Clausal negation vs. negative focus constructions

According to [3, 4], focus-sensitive expressions can be divided into (at least) two separate classes: (i) those that exhibit non-conventionalized focus sensitivity (e.g., *always* and counterfactual conditionals); and (ii) those whose focus sensitivity is conventionalized and lexically encoded (e.g., *only* and *even*). Clausal negation constructions such as those in (5) belong to the former category. They are focus sensitive in that minimal pairs differing in focus marking produce different inferences. For instance, *I didn't invite* JOHN *for dinner* may suggest that the speaker invited someone else for dinner, while *I didn't invite John for* DINNER may suggest that the speaker invited John to another event. However, these inferences are cancelable implicatures rather than entailments (see [8] for a recent review). For example, a sentence like *I obviously can't afford a* MERCEDES may imply that the speaker can afford a cheaper car, but it can just as well be coherently followed by the claim *In fact, I can't afford any car!*

(5) a. The cookies did not all burn. (floating quantifier construction)
b. I didn't burn all of the cookies. (QP in object position)

Conventionalized focus-sensitive expressions, by contrast, must associate with focus, and they

encode a dependency on focus marking [3, 4]. I argue that NFCs such as those in (6) belong to this category. This is fairly uncontroversial as far as negated clefts (6b) are concerned (see e.g., [18]). I assume that negated quantifier phrase constructions such as (6a) are NFCs too, and their semantics is the same as negated clefts. This assumption is motivated in Section 3.

(6) a. Not all of the cookies burned. (negated quantifier phrase construction)
b. It's not every student who has access to their teacher 24/7.[1] (negated cleft)

Further evidence for the contrast between clausal negation and NFCs comes from Hebrew and Korean. In Hebrew, negated quantifier phrase constructions, as in (7b), contrast with various clausal negation constructions. For instance, (7a) involves clausal negation with VS word order and lacks an SI in this scenario. The NFC in (7b), on the other hand, triggers a 'not all but some' SI even though it contradicts the CG given that the cookies weren't even baked.

(7) *Context: John was baking cookies. He put the cookies in the oven, forgetting to set up an alarm. Two hours later, he realized that he forgot about the cookies. He rushed to the kitchen and opened the oven only to learn that he also forgot to turn on the oven.*
a. rak be-mazal lo nisrefu l-i kol ha-ugiyot.
only in-luck NEG burned to-me all the-cookies
'It was only through luck that the cookies did not all burn (possibly none did).'
b. #rak be-mazal lo kol ha-ugiyot nisrefu l-i.
only in-luck NEG all the-cookies burned to-me
'It was only through luck that not all of the cookies burned (but some did).'

Korean presents an interesting example because it usually employs another type of NFC to express the meaning of 'not all' (Korean lacks negated quantifier phrase constructions). In (8), the contrastive marker *nun* is attached to the lexical verb, and the QP is focused (the particle *nun* requires a focused constituent in its scope). In accordance with the generalization, (8) triggers an SI even in the context of (1), where the SI contradicts the CG (Eunsun Jou, p.c.).

(8) Nehuy motwu-ka nakceyha-ci-nun anh-un sasil-i nollap-ta.
you all-NOM fail-NMLZ-CNTR NEG-ADN fact-NOM surprising-DECL
'(lit.) The fact that not all of you failed is surprising.'

3 Analysis

I argue that the contrast between clausal negation constructions and NFCs is due to the fact that their SIs come from different sources. This difference cannot be detected in sentences where 'not all' is unembedded, as in (5) and (6), because the sentences in (5) give rise to robust SIs too. However, if 'not all' is embedded in a (factive or non-factive) SDE context, the difference between the two types of constructions comes to the surface. In clausal negation constructions such as (5), the SI 'not all but some' is a cancelable implicature, which is derived by general principles of implicature calculation (whether one assumes a pragmatic or grammatical approach to scalar implicatures). Therefore, these 'not all' constructions pattern with other scalar expressions such as 'or' in that the presence of an SI is affected by the monotonicity of the environment in which the scalar expression is embedded.

On the other hand, I propose that implicature calculation is not responsible for the SIs in NFCs such as (6). Of course, these SIs could be derived by implicature calculation, but there

[1] For some reason, *every* is much more frequent than *all* in negated clefts, which is why I use it here.

is also another way to derive them—namely, from the existential presupposition triggered by the NFC. For ease of presentation, I will start with negated clefts (6b). I follow the analysis of clefts in [18], who propose that the existential presupposition comes from the question under discussion (QUD) that the cleft addresses. In (6b), the QP (or just the quantifier) is focused. The QUD is (9a), and it contains the alternatives in (9b). Accepting a QUD introduces a presupposition that one of the propositions in the alternative set is true [14]. Therefore, the cleft presupposes that the disjunction of (9b) is true, which gives rise to the existential presupposition in (9c). If we combine the assertion (9d) with the existential presupposition (9c), we get a 'not all but some' SI. Importantly, this SI is derived independently of whatever general mechanism one assumes for implicature calculation. Given that an SI triggered by an NFC does not exhibit the characteristic context dependence of implicatures, I propose that the existential presupposition is indeed the source of the SI. Due to projection, the SI is present even in non-factive SDE contexts, as in (2b), and when it contradicts the CG, as in (1b) and (7b).[2]

(9) a. QUD: How many students have access to their teacher 24/7?
b. Focus alternatives: {Some/Most/All students have access to their teacher 24/7.}
c. Presupposition: Some students have access to their teacher 24/7.
d. Assertion: The alternative with *all* is rejected (which is a partial answer to the QUD) ≈ 'It is not the case that every student has access to their teacher 24/7.'

I argue that negated quantifier phrase constructions such as (6a) are NFCs too, and their semantics is the same as negated clefts (although they have a different distribution in English). This relates to a recent debate about the syntactic structure of these constructions. [5] suggests that negation modifies the entire QP, and the resulting constituent fills the subject position, as in (10a). On the other hand, [7] argue that negation is in a high position in the left periphery, and the QP is adjacent to it because it is fronted to a focus position, as in (10b). The main evidence for the latter structure comes from clauses in which the QP functions as the complement of a preposition. In Spanish, the preferred word order is one in which negation precedes the PP (i.e., Neg + Prep + QP), while the alternative order in which the preposition precedes the negated quantifier phrase (i.e., Prep + Neg + QP) is rejected by some speakers and is not available at all for certain quantifiers [7]. This preference is not particular to Spanish but found across a wide variety of languages [1], e.g., in Hebrew (11), where the alternative word order is not possible. This suggests that negation does not form a constituent with the QP. Instead, the QP (or PP containing it) is fronted. If this analysis is correct, one expects (10b) to be a conventionalized NFC because it necessarily associates with focus, like a negated cleft.

(10) a. $[_{CP}\ [_{IP}\ [_{NegP}$ **Not** $[_{QP}$ **all** of the students]] came]].
b. $[_{CP}\ [_{NegP}$ **Not** $[_{FocP}\ [_{QP}$ **all** of the students$]_1$ Foc^o] $[_{IP}\ t_1$ came]]].

(11) lo al kol mucar mecuyan ha-mexir.
NEG on every product is.marked the-price
'Not every product has a price tag on it.'

The presuppositional status of the SIs triggered by NFCs is evidenced by projection facts. In (12a), the SI projects from the scope of a probability adverb, which is characteristic of presuppositions. In contrast, regular implicatures typically do not arise in this context (12b). Another example is given in (13), where 'not all' occurs in the complement of the modal predicate *hope*. SIs are usually not generated under *hope* (e.g., *Mary hopes to get into Harvard or Yale* $\not\rightsquigarrow$ *Mary doesn't wish to get into both Harvard and Yale*). This is indeed what we find

[2] Regarding (2b), see [15] for other cases where presuppositions interact with the semantics of counterfactuals.

in (13a)—i.e., we infer that the speaker probably hopes that no audience member will throw tomatoes at him and his fellow actors. In contrast, (13b) conveys that the speaker accepts his fate of being pelted with tomatoes by some audience members but hopes that others will not take part in the tossing. Thus, the SI projects from the modal context in (13b).

(12) a. It's very likely that not all of the cookies burned. ⇝ Some of the cookies burned.
b. It's very likely that some of the cookies burned. ⇝̸ Not all of the cookies burned.

(13) *Context: After a terrible theatrical performance, the actors are about to take their bows. One of the actors says to his friends:*
a. I hope they don't all start throwing tomatoes at us!
b. I hope not all of them start throwing tomatoes at us!

4 Other scalar expressions in factive SDE contexts

In this section, I briefly discuss the behavior of other scalar expressions in factive SDE contexts. First, I demonstrate that in contrast to previous claims in the literature, SIs are merely optional in this environment, while noting that the scalar item 'some' is somewhat exceptional in this respect. Second, I maintain that despite appearances, the SI triggered by 'some' is not of the same nature as the one triggered by 'not all' in NFCs. Finally, I argue that these observations pose a challenge for several previous accounts.

It has been argued that a sentence like (14a), where 'some' occurs in a factive SDE context, is infelicitous in the context of (1) due to competition with the alternative with 'all' (14b) which has a stronger presupposition entailed by the CG [9, 16]. Thus, (14a) can be said to trigger an obligatory SI in this context since it presupposes that not all of the students passed.

(14) a. #Frankly, I was surprised that SOME of you passed! (uttered in the context of (1))
b. Frankly, I was surprised that all of you passed!

However, I observe that this pattern is idiosyncratic to the scalar item 'some'. We have already seen that 'not all' (with clausal negation) and 'or' do not always give rise to SIs in factive SDE contexts, e.g., (1a) and (4), respectively. Another example is that if we replace *some* with *any* in (14a), the resulting sentence is felicitous and lacks an SI (15). According to common assumptions (see e.g., [10]), *any* is an existential like *some*, and indeed, outside the context of (1), (15) would probably give rise to a 'some but not all' SI just like (14a). Yet, (15) shows that this SI is not obligatory in any way. A final example is with *allow*, which sometimes gives rise to an SI in factive SDE contexts (16a), but not always (16b). I conclude from these examples that the obligatory SI of 'some' in cases like (14a) is in fact the exception rather than the rule.

(15) Frankly, I was surprised that ANY of you passed! (uttered in the context of (1))

(16) a. John is unaware that the students are allowed to take a bowling class for credit.
⇝ The students are allowed, but not required, to take a bowling class for credit.
b. A: As part of your studies, you are required to take a bowling class for credit.
B: Required?! I'm surprised I am ALLOWED to!

The question of why 'some' differs from other scalar expressions is beyond the scope of this paper. However, I will make a few relevant observations in order to clarify the difference between the infelicity of (14a) and (1b). While these examples appear to be parallel, the parallelism breaks down if the matrix clause is negated: (17a) becomes felicitous (i.e., 'some' does not trigger an SI), whereas (17b) is still bad due to a 'not all but some' SI. I speculate that (14a)

is infelicitous because 'some' is not a proper scalar endpoint. That is, the intended meaning of (14a) is that Mary would have been surprised even if only one student had passed. This is conveyed by (15) with *any*, while (14a) with *some* strongly implies that more than one student passed. Hence, the oddness of (14a) appears to be related to the fact that speakers consider *some* unnatural for reference to a single entity [6]. The CG in (17), by contrast, favors a 'more than one' interpretation, so (17a) is felicitous (whereas a variant with *any* would not be). This suggests that the SIs of (14a) and (1b) come from different sources (see Section 3).

(17) *Context: Mary expected that 40-70% of the students would pass. Surprisingly, all failed.*
   a. I wasn't surprised that SOME of you failed, but I certainly did not expect that ALL of you would!
   b. #I wasn't surprised that not ALL of you passed, but I certainly did not expect that NONE of you would!

The fact that SIs are typically optional in SDE contexts poses a challenge for several previous studies. [16] propose that a sentence is infelicitous if there is a relevant alternative with a stronger presupposition which is entailed by the CG. This principle predicts the infelicity of 'some' in (14a), but it seems to make incorrect predictions regarding the other scalar expressions discussed above, namely 'not all' (1a), 'or' (4), 'any' (15), and 'allow' (16b).

Another approach suggests that (14a) is infelicitous because of obligatory exhaustification [13, 2]. According to [11, 12], a silent EXH operator is obligatorily adjoined to any clause. Additionally, if an alternative is contextually equivalent to the prejacent of EXH, it is necessarily relevant and must be excluded by EXH. In (18) (from [12]), the logically weak alternative (18a) is contextually equivalent to the strong alternative (18b) given that all of the students get the same grade. According to this analysis, (18a) is odd because EXH must exclude the strong alternative, and the resulting SI contradicts the CG.

(18) *Context: Prof. Smith always assigns the same grade to all of the students in his class.*
   a. #This year, Prof. Smith assigned an A to some of his students.
   b. This year, Prof. Smith assigned an A to all of his students.

The example with 'some' (14a) and the alternative with 'all' (14b) are not contextually equivalent in the context of (1) because (14a) makes a stronger assertion than (14b) [16]. However, the presupposition of (14b) is trivially relevant because it is entailed by the CG [13]. Therefore, this alternative is excluded by EXH, which results in a SI that contradicts the CG [13, 2]. Notice that this account, too, appears to wrongly predict (1a) to be infelicitous by the same reasoning that was applied to (14a): the prejacent of EXH with 'not all' has a relevant alternative with 'not some' (i.e., 'none'), and excluding the latter would lead to conflict with the CG. A final example is given in (19), which is modeled after (18). Assume an LF with embedded EXH (19a), as in [2]. The presupposition of the alternative with 'not any' (19b) is relevant because it is entailed by the CG (and contextually equivalent to the presupposition of (19a)), and hence must be excluded by EXH. Yet, (19a) is fine and does not presuppose that some of the students got an F. This shows that the alternative with 'not any' is in fact not excluded.

(19) *Context: This year, all of Prof. Smith's students did very poorly except for Jane, whose work well deserved an A. He reasons that it is better that ten bad students pass than that one good student fail, and decides to give all the students a C. He then tells them:*
   a. It is only because of Jane that [EXH [I didn't give all of you an F]].
   b. Relevant alternative: It is only because of Jane that I didn't give any of you an F.

References

[1] Omri Amiraz. *The scope of negation*. PhD thesis, The Hebrew University of Jerusalem, Jerusalem, forthcoming.

[2] Itai Bassi, Guillermo Del Pinal, and Uli Sauerland. Presuppositional exhaustification. *Semantics and Pragmatics*, 14, 2021.

[3] David I Beaver and Brady Z Clark. *Always* and *only*: Why not all focus-sensitive operators are alike. *Natural Language Semantics*, 11(4):323–362, 2003.

[4] David I Beaver and Brady Z Clark. *Sense and sensitivity: How focus determines meaning*. John Wiley & Sons, 2009.

[5] Chris Collins. Outer negation of universal quantifier phrases. *Linguistics and Philosophy*, 43(3):233–246, 2020.

[6] Judith Degen and Michael K Tanenhaus. Processing scalar implicature: A constraint-based approach. *Cognitive Science*, 39(4):667–710, 2015.

[7] Ricardo Etxepare and Myriam Uribe-Etxebarria. Crosslinguistic variation in constituent negation. In *Footprints of phrase structure*. John Benjamins, Amsterdam, forthcoming.

[8] Anamaria Fălăuş. Negation and alternatives: Interaction with focus constituents. In Viviane Déprez and M Teresa Espinal, editors, *The Oxford Handbook of Negation*, pages 333—348. Oxford University Press, Oxford, 2020.

[9] Jon Gajewski and Yael Sharvit. In defense of the grammatical approach to local implicatures. *Natural Language Semantics*, 20(1):31–57, 2012.

[10] William A Ladusaw. *Polarity sensitivity as inherent scope relations*. Garland Pub, New York, 1980.

[11] Giorgio Magri. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17(3):245–297, 2009.

[12] Giorgio Magri. Another argument for embedded scalar implicatures based on oddness in downward entailing environments. *Semantics and Pragmatics*, 4:1–51, 2011.

[13] Paul Marty and Jacopo Romoli. Presuppositions, implicatures, and contextual equivalence. *Natural Language Semantics*, 29(2):229–280, 2021.

[14] Craige Roberts. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5:1–69, 1996.

[15] Mats Rooth. Association with presupposition? In Peter Bosch and Rob A van der Sandt, editors, *Focus: Linguistic, cognitive, and computational perspectives*, pages 232–246. Cambridge University Press, Cambridge, 1999.

[16] Benjamin Spector and Yasutada Sudo. Presupposed ignorance and exhaustification: How scalar implicatures and presuppositions interact. *Linguistics and Philosophy*, 40(5):473–517, 2017.

[17] Bob van Tiel, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. Scalar diversity. *Journal of Semantics*, 33(1):137–175, 2016.

[18] Dan Velleman, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea, and Elizabeth Coppock. It-clefts are IT (inquiry terminating) constructions. In Anca Chereches, editor, *Proceedings of SALT 22*, pages 441–460. LSA, Washington DC, 2012.

[19] Kai von Fintel. NPI licensing, Strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999.

Disentangling Semantic Empty-Set Effects in Quantifier Comprehension From Simple Associations: Processing Evidence From Exceptive-Additives

Nadine Balbach[1]*, Fabian Schlotterbeck[1]*, Hening Wang[1], and Oliver Bott[2]

[1] University of Tübingen, Tübingen, Germany
{nadine.balbach,fabian.schlotterbeck,hening.wang}@uni-tuebingen.de
[2] University of Bielefeld, Bielefeld, Germany
oliver.bott@uni-bielefeld.de
*Co-first authors.

Abstract

According to a recent proposal, so-called empty-set (ES) quantifiers are extraordinarily difficult to interpret because their verification algorithm involves one specific rule intended to handle ES situations. The rule relies on the explicit representation of objects in the restrictor set that are not in the scope set. This type of encoding is assumed to be cognitively costly because the cognitive system is tuned to encoding the presence of properties rather than their absence, a strategy viable for non-ES quantifiers but doom to failure for ES quantifiers. While previous data support this assumption, they do not rule out the possibility that the observed ES effects were at least in part due to a superficial mismatch between the visual contexts and the content words in the tested sentences. We report on a picture verification experiment in which we combined ES quantifiers with German *außer*-('besides')-phrases to investigate how the additive inference, triggered by *außer*, interacts with ES effects. Our results indicate that ES effects are in fact based on compositional meaning and cannot be explained based on superficial features of the stimuli alone.

1 Introduction

Recently, [5] proposed an algorithmic theory of quantifier interpretation in natural language based on witness sets in order to account for a number of psycholinguistic findings. One specific distinguishing feature of this theory, e.g. in comparison to Generalized Quantifier Theory [20] and algorithmic implementations thereof in terms of semantic automata [4, 24], is that it predicts so-called empty-set (ES) quantifiers, which have the empty set as a witness set (i.e. they are true in a situation in which none of the restrictor elements is also in the scope set (ES situation)), to be more difficult to interpret than quantifiers that do not have the ES among their witnesses. Since all downward entailing quantifiers are ES quantifiers, the proposal of [5] explains the well-known processing difficulty associated with downward entailing quantifiers (e.g. [14, 18, 7, 2]). What's more, the proposal also predicts extraordinary difficulty (the so-called ES effect) in case ES quantifiers are verified in ES situations. In a series of combined self-paced reading plus picture verification experiments, [5] confirmed these predictions across a range of different types of quantifiers. Moreover, they teased apart the difficulty of ES quantifiers from effects of monotonicity by establishing ES effects for the non-monotone empty-set quantifier *none or three*. In the current study, we combined ES quantifiers with the German so-called exceptive-additive *außer* ('besides', cf. [27, 28]) to investigate how the additive inference, triggered by *außer*, interacts with ES effects. This allowed us to further test the algorithmic theory of [5] by ruling out an alternative explanation of ES effects that is based on a superficial mismatch between linguistic and visual features.

2 ES effects and their explanation

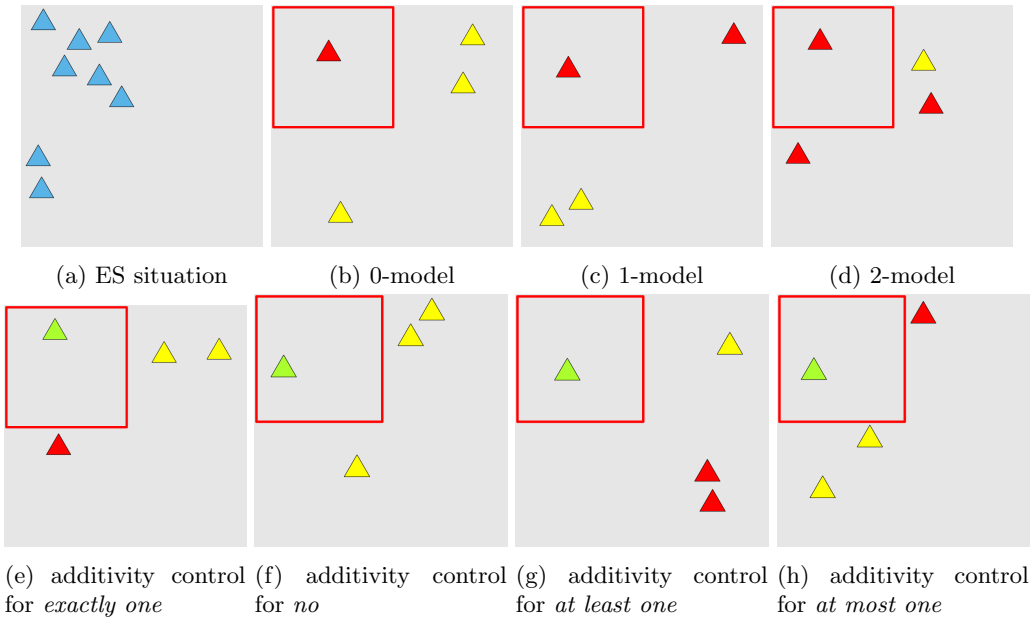
According to [5], ES quantifiers are difficult because their verification algorithm involves one specific rule intended to handle ES situations. This rule relies on the explicit representation of objects in the restrictor set that are not in the scope. [5] argue that this type of encoding is cognitively costly because the cognitive system is tuned to encode the presence of properties rather than their absence, a strategy viable for non-ES quantifiers but doomed to failure for ES quantifiers. While the data of [5] support the assumption that the verification algorithm of ES quantifiers is costlier than the one of non-ES quantifiers, they do not rule out the possibility that the observed ES effects were at least in part due to a superficial mismatch between the visual contexts and the content words in the tested sentences. To illustrate this explanation, consider a sentence like *at most one triangle is red* in the context of a picture like in Fig 1a (below). The sentence is true although none of the objects has the mentioned color (i.e. none is red). The lexical meaning of the adjective *red* does not match any of the color features of the given visual context. There is thus a conflict between the truth-value and the semantic dis-similarity between visual features and content words. This mismatch in ES situations is restricted to ES quantifiers and could thus explain at least some part of the observed ES effects.

3 Testing empty-set effects in *außer* ('besides') sentences

Modifying ES quantifiers with the so-called exceptive-additive *außer* ('besides') (cf. [27, 28]), allows for testing ES effects in the absence of the type of superficial mismatch just described.This is because an *außer*-phrase can explicitly mark the properties of some objects whose presence are a prerequisite for computing the truth-value but are not necessarily part of the asserted content. As a consequence, it is possible to create ES-situations without a visual mismatch between the predication and the picture. For a sentence such as (1), this is illustrated by the model in Fig. 1b, which constitutes an empty-set situation even though there are red triangles in the picture. The truth-value is determined by evaluating the quantifier, Q, with respect to all triangles except the marked one and the prerequisite is that the marked triangle is in the scope set of the quantifier (i.e. the red objects). In the present paper, we call the former meaning component the *quantificational claim* and the latter the *additive inference* to set them apart from [19]'s generality and exception claims.

(1) Außer dem markierten | ist | Q | weiteres | Dreieck | rot.
Besides the marked | is | Q | more | triangle | red.
'Besides the marked one, there are Q more triangles that are red.'

Quantifiers combined with *außer*-phrases provide also an interesting test case for the present follow-up study in terms of their inherent constructional quantificational complexity relative to the cases already investigated in [5]. Bott et al. compared ES effects in simply quantified sentences with those in doubly quantified sentences forming iterations of quantifiers [22] and observed ES effects of different strengths in picture verification. ES effects varied between less than 30% errors for ES quantifiers in simply quantified sentences as compared to more than 50% errors for ES quantifiers in doubly quantified sentences. ES quantifiers in combination with exceptive-additive phrases are therefore of additional interest from a theoretical perspective because they involve relations between three sets of entities, i.e. they are akin to quantifiers of type $\langle 1,1,1\rangle$ (modulo non-assertive status of some aspects of their meaning, cf. [19]). Considering the semantic types of the relations involved, they thus fall in between sentences with a single type $\langle 1,1\rangle$ quantifier and iterations of such quantifiers, resulting in type $\langle 1,1,2\rangle$.



(a) ES situation (b) 0-model (c) 1-model (d) 2-model

(e) additivity control for *exactly one* (f) additivity control for *no* (g) additivity control for *at least one* (h) additivity control for *at most one*

Figure 1: Example visual stimuli

Außer ('besides') constitutes a prime example for the postulated class of exceptive-additives. In combination with a restricted class of quantifiers (called generalizers in [19]), it can nevertheless also receive the better studied 'pure exceptive reading' (for combinatory restrictions and the analysis of 'pure exceptives', see e.g. [19, 12, 10, 9, 25, 15, 13]). In combination with the quantifiers tested in the present study (see below), *kein* ('no') is the only one that allows for a 'pure exceptive reading'. However, the additive inference is assumed to remain the same even in these cases. The only difference may consist in the fact that [27, 28] considers the inference to be a presupposition whereas others remain agnostic in this regard (see e.g. [19]). The present study, therefore, included a control experiment in addition to the main experiment testing for the additive inferences and their semantic status, that is whether the quantifiers under investigation give rise to an additive inference and whether this inference should be considered as asserted or presupposed. Last but not least, neither 'pure' exceptives nor exceptive-additives have received much attention in experimental work (for exceptives see [16, 17]) and are therefore interesting in their own right.

4 Methods

We adopted the methods from [5] and conducted a combined self-paced reading plus picture verification experiment (80 native German speakers, all self-reported students, 41 female, 39 male, recruited via Prolific). 4 of these participants were excluded from the data analysis due to fails in the attention checks or due to too many timeouts.

In the web-based experiment participants first received German sentences as in (1) employing a self-paced reading paradigm with moving window presentation (phrase by phrase, as indicated by vertical lines). After having read the final region of the sentence, the sentence disappeared and instead, a picture as in Fig. 1b-1d was shown in the middle of the screen for which a truth-

value judgment had to be provided within a time limit of 10 s. In the verification stage, both judgments and judgment RTs were recorded. Reading times and picture verification latencies were preprocessed by adopting the methods from [5, Exp. 2].

As for the sentence materials, quantifiers were manipulated in a 2×2 design. We compared degree quantifiers (*höchstens ein* ('at most one') and *mindestens ein* ('at least one')) to non-degree quantifiers (*kein* ('no') and *genau ein* ('exactly one')) and in both classes compared an ES quantifier (*höchstens ein* and *kein*, respectively) to an otherwise similar non-ES quantifier. The four sentences with quantifiers were accompanied each by a picture showing one marked object in the mentioned color and either zero, one, or two additional objects in that same color plus objects in another color, resulting in 12 experimental conditions in total. In addition, the four sentences were also combined with the pictures in Fig. 1e-1h. These control conditions violate the additive inference of *außer*. In the models of all of these four conditions the additive inference was falsified because the marked object was never in the mentioned color whereas the rest of the objects were colored in such a way to make the quantificational claim true. In total, 64 items were created in the 16 conditions each and were distributed together with 84 fillers across 16 lists using a Latin Square design.

5 Results

Prior to further analysis, reading times below 200 ms and above 2.5 standard deviations above a participant's mean RT were removed (removal of 0.2% of the data).[1] Furthermore, judgment RTs were corrected for differences between response types (expected answer: TRUE vs. FALSE) by computing the fixed effect (-154 ms) as well as by-participants random intercepts of response type (cf. [1]) in a linear mixed effects regression (LMER, [3]) on a set of designated filler items and subtracting these effects from the corresponding RT in the main experiment. Below, the resulting residual judgment RTs are reported. Throughout the study, we analyzed log judgment RTs using LMER and the accuracy of judgments using generalized (logit) mixed effects regression (GLMER, [3]) modeling. In the statistical analysis of the accuracy data, we aggregated responses across 1- and 2-models. The reason was that a comparison between 1- and 2-models is problematic because correct responses in these conditions may differ between quantifiers and, in contrast to RTs, responses themselves cannot easily be corrected for response bias, etc. For the random effects structure, we started with including by-participants and by-items random intercepts and then included random slopes for all the relevant fixed factors until the models did not converge or no longer improved model fit.

Judgment RTs and accuracy across the twelve target conditions are shown in Fig. 2. In the target conditions, log-likelihood ratio tests revealed a significant three-way interaction between QUANTIFIER TYPE, ES PROPERTY and MODEL ($\chi^2(2) = 6.31, p = .042$). This was due to qualitatively similar but quantitatively more pronounced effects in *degree* as compared to *non-degree* quantifiers: In both QUANTIFIER TYPES, we found the predicted two-way interaction between ES PROPERTY and MODEL (*degree*: $\chi^2(2) = 53.09, p < .001$; *non-degree*: $\chi^2(2) = 18.22, p < .001$). The reason for these interactions was that, for non-ES quantifiers, additional objects in the mentioned color (cf. Fig. 1b-1d) led to an increase in RT whereas for ES quantifiers the ES situation caused deviation from this pattern and led to the longest RTs (0- vs. 1-model in *degree* ES quantifiers: $t = 6.06, p < .001$; and in *non-degree* ES quantifiers: $t = 3.94, p < .001$). In the accuracy, we observed an ES effect for the *degree* ES quantifier (*at most one*), which led to only 54.7% correct responses. All the other quantifiers had accuracy above 90% in all models.

[1] Data and analysis scripts are provided at https://osf.io/w8d95/.

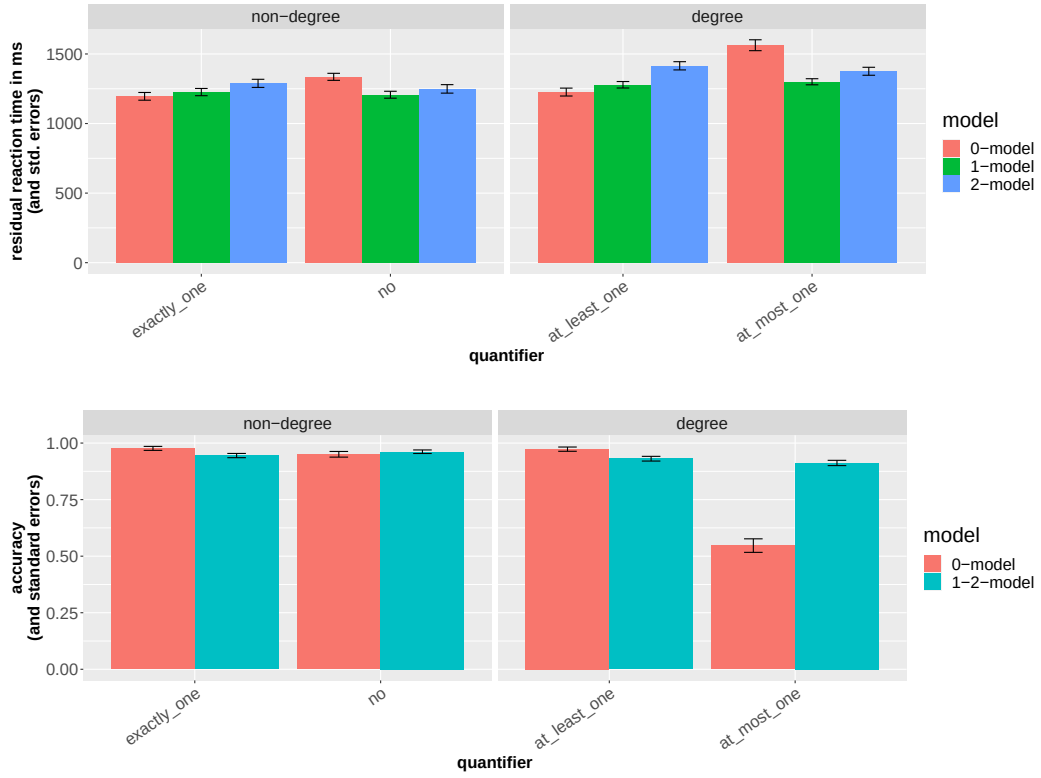


Figure 2: Verification data from target conditions in the experiment. Top: RT; bottom: accuracy.

This pattern in accuracy also led to a significant three-way interaction in a GLMER analysis with the same fixed effects and the same random effects structure as for the LMER analysis ($\chi^2(1) = 8.25, p = .004$). The three-way interaction was due to the fact that accuracy was significantly lower in the 0-model compared to the other two types of models for *at most one* ($\beta = -1.32, z = -10.6, p < .001$), not significantly different for *no* ($\beta = -0.15, z = -0.84, p = .4$), but significantly higher for the other two quantifiers ($\beta = 0.47, z = 2.43, p = .015; \beta = 0.43, z = 2.11, p = .035$). Taken together, *at most one* thus led to rather robust empty-set effects both in judgments and judgment RTs, whereas ES effects for *no* turned out to be weaker affecting only judgment RTs but not the judgments themselves.

The violation of the additive inference in the controls slowed down the verification process and decreased accuracy: Residual RTs for control conditions were between approximately 1,500 and 1,650 ms, comparable to the longest RTs among the target conditions, but there were no significant differences within the control conditions ($\chi^2(3) = 6.13, p = .105$). Similarly, accuracy was between 47% and 61 % and thus in the range of the lowest accuracy among the target conditions. The evaluation of sentences with *kein* was easier than with other quantifiers: A GLMER analysis with *kein* as reference level revealed significant differences between *kein* and all the other quantifiers (*at most one*: $\beta = -1.16, z = -3.87, p < .001$; *at least one*: $\beta = -0.87, z = -3.39, p < .001$; *exactly one*: $\beta = -0.70, z = -2.16, p = .031$).

6 Discussion

We take the long RT and neither clearly true nor false judgments from the four control conditions as evidence that the tested *außer*-sentences do, as intended, trigger an additive inference (e.g. that the marked object is also red) that is, however, not part of the asserted meaning but rather non-at-issue or presupposed content (cf. [27, 28]). In the twelve experimental conditions, we found ES effects for both *degree* and *non-degree* quantifiers. Because the marked object always had the color mentioned in the sentence, these results indicate that ES effects are, in fact, related to compositional meaning and cannot be explained based on superficial features of the stimuli alone. The finding that ES effects were larger within degree than within non-degree quantifiers is consistent with the results of [5], who also found less pronounced ES effects for the *non-degree* quantifier *kein* ('no') as compared to modified numerals. This may be related to the fact that *kein* is the simplest possible ES quantifier in the sense that its verification needs fewer operational steps than other ES quantifiers like, e.g., the superlative *degree* quantifier *höchstens ein* ('at most one') (see also [11] on the complexity of *at most*). A direct comparison of effects across experiments is, however, problematic, especially since [5] tested *kein* in a Boolean combination like *none or three*. Nevertheless, another parallel between the present study and [5] is noteworthy: The extremely high error rate of almost 50% for the superlative ES quantifier *höchstens ein* that we observed in the present experiment resembles an accuracy even below the 50% chance level that [5] found in doubly quantified sentences containing *at most*. This parallel may point to an interaction between compositional structure of a sentence as encoded in semantic types and the cognitive complexity of the individual quantifiers it contains, specifically regarding the ES property. As pointed out in section 3, the complexity of *außer*-sentences in terms of their semantic type lies between the complexity of sentences with one single quantifier and of iterations of two quantifiers in a doubly quantified sentence. Further experiments are required to directly test for this type of interaction between different sources of complexity.

Apart from indicating an additive inference, data from the control conditions indicate that participants were uncertain how to judge violations of this inference. Assuming a presuppositional status of the inference (as, e.g., in [27, 28]), these results are consistent with the assumption that a sentence whose presupposition is not met is neither TRUE nor FALSE (see e.g. [23, 8, 26]) and with previous experimental studies that looked at falsified presuppositions (see [29, 6, 21] but also [1]). The relatively high accuracy of *kein* ('no') as compared to the other three controls may be related to the fact that this was the only quantifier among those we tested that can uncontroversially also be interpreted as a 'true' exceptive (cf. [19]), albeit with the same truth-conditional import. Alternatively, it may be explained by the fact that this was the only condition among the controls for which the quantificational claim involved a potential ES effect.

In our view, exhausting the whole compositional potential of combining domain subtraction with the ES property is theoretically challenging and empirically promising at the same time. If we put ES quantifiers inside the *außer*-phrase as in *all but at most five*, we can push the distinction between lexical features and truth-values to the extreme and study 'embedded ES effects' in situations where all elements from the restriction are also in the scope set. At the same time, *all but at most five* is an upward entailing ES quantifier that allows for 'global ES effects' in situations where the restrictor set contains fewer than five elements. In particular, this is the case if none of those elements is in the scope set. ES effects of that kind would allow us to follow up on one of the goals of [5] by completely separating monotonicity from the ES property.

Acknowledgments

We thank our student assistants Sarah Ackerschewski, Julia Nieslony, Tim Tichy, and Margarethe van Liempt for helping with the implementation of the experiment. Furthermore, we would like to thank the anonymous reviewers. A special thanks goes to Ekaterina Vostrikova for valuable feedback and discussion on exceptive-additives. Nadine Balbach received funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 833 – Project ID 75650358.

References

[1] Márta Abrusán and Kriszta Szendrői. Experimenting with the king of France: Topics, verifiability and definite descriptions. *Semantics & Pragmatics*, 6:1–43, 2013.

[2] Petra Augurzky, Fabian Schlotterbeck, and Rolf Ulrich. Most (but not all) quantifiers are interpreted immediately in visual context. *Language, Cognition and Neuroscience*, 35(9):1203–1222, 2020.

[3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[4] Johan van Benthem. *Essays in Logical Semantics*, volume 29 of *SLAP*. Springer, Dordrecht, 1986.

[5] Oliver Bott, Fabian Schlotterbeck, and Udo Klein. Empty-set effects in quantifier interpretation. *Journal of Semantics*, 36(1):99–163, 2019.

[6] Emmanuel Chemla and Lewis Bott. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and cognitive processes*, 28(3):241–260, 2013.

[7] Isabelle Deschamps, Galit Agmon, Yonatan Loewenstein, and Yosef Grodzinsky. The processing of polar quantifiers and numerosity perception. *Cognition*, 143:115–128, 2015.

[8] Gottlob Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100(1):25–50, 1892.

[9] Jon Gajewski. Npi any and connected exceptive phrases. *Natural Language Semantics*, 16(1):69–110, 2008.

[10] Iván García Álvarez. *Generality and exception: A study in the semantics of exceptives*. PhD thesis, PhD dissertation, Stanford University, 2009.

[11] Bart Geurts, Napoleon Katsos, Chris Cummins, Jonas Moons, and Leo Noordman. Scalar quantifiers: Logic, acquisition, and processing. *Language and Cognitive Processes*, 25:130–148, 2010.

[12] Aron Hirsch. An unexceptional semantics for expressions of exception. *University of Pennsylvania Working Papers in Linguistics*, 22(1):16, 2016.

[13] Jack Hoeksema. The semantics of exception phrases. *Quantifiers, logic, and language*, pages 145–177, 1995.

[14] Marcel Adam Just and Patricia Ann Carpenter. Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behaviour*, 10(3):244–253, 1971.

[15] Friederike Moltmann. Exception sentences and polyadic quantification. *Linguistics and Philosophy*, 18(3):223–280, 1995.

[16] Prerna Nadathur and Dan Lassiter. An experimental look at the negative implications of exceptives. In Sherry Hucklebridge and Max Nelson, editors, *Proceedings of NELS 28*, vol. 2, pages 201–210. GLSA, 2018.

[17] Prerna Nadathur and Daniel Lassiter. Unless: An experimental approach. In Eva Csipak and Hedde Zeijlstra, editors, *Proceedings of Sinn und Bedeutung 19*, page 446–464, 2015.

[18] Mante S. Nieuwland. Quantification, prediction, and the online impact of sentence truth-value: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory,*

and Cognition, 42(2), 2016.

[19] Stanley Peters and Dag Westerståhl. The semantics of exceptives. *Linguistics and Philosophy*, Published Online, 07/08/2022.

[20] Stanley Peters and Dag Westerståhl. *Quantifiers in Language and Logic*. Oxford University Press, New York, 2006.

[21] Florian Schwarz. False but slow: Evaluating statements with non-referring definites. *Journal of Semantics*, 33:177–214, 2016.

[22] Shane Steinert-Threlkeld and Thomas F. Icard III. Iterating semantic automata. *Linguistics and Philosophy*, 36:151–173, 2013.

[23] Peter F. Strawson. On referring. *Mind*, 59(235):320–344, 1950.

[24] Jakub Szymanik. *Quantifiers and Cognition: Logical and Computational Perspectives*, volume 96 of *SLAP*. Springer, Dordrecht, 2016.

[25] Kai Von Fintel. Exceptive constructions. *Natural language semantics*, 1(2):123–148, 1993.

[26] Kai von Fintel. Would you believe it? The king of France is back! (Presuppositions and truth-value intuitions). In Marga Reimer and Anne Bezuidenhout, editors, *Descriptions and Beyond*. Oxford University Press, 2004.

[27] Ekaterina Vostrikova. Besides exceptives. In Duygu Göoksu Maggie Baird and Max Nelson, editors, *Proceedings of NELS 49*, pages 265–274, 2019.

[28] Ekaterina Vostrikova. *Phrasal and clausal exceptive-additive constructions crosslinguistically*. PhD thesis, UMass Amherst, 2019.

[29] Jérémy Zehr. *Vagueness, presupposition and truth-value judgments*. PhD thesis, École Normale Supérieure de Paris, 2014.

Positivity, (anti-)exhaustivity and stability*

Moshe E. Bar-Lev and Roni Katzir

Tel Aviv University
moshebl@tauex.tau.ac.il
rkatzir@tauex.tau.ac.il

Abstract

Why are AND and OR the only binary connectives which are lexicalized as simplex? In this paper we point out a special property they have: They are the only 'meaningful' connectives which enable the innocent exclusion of alternatives that are logically independent from the connective's arguments. We further observe that, as a result, they are the only meaningful connectives that can block prior probabilities from leading to anti-exhaustive interpretations (Cremers et al. 2022). This property of AND and OR can explain their lexicalizability once combined with the notion of Stability from Bar-Lev & Katzir (2022), while the theory that emerges has advantages over the Stability-based account of the typology we proposed in Bar-Lev & Katzir (2022).

1 Positivity

The typology of logical connectives presents a major cross-linguistic puzzle: Among the 12 binary logical connectives in table 1 (considering here only those that are 'meaningful', i.e., non-trivial and non-redundant),[1] only AND and OR are ever lexicalized as simplex. While NOR is also apparently lexicalized sometimes, it has been argued that whenever it is, it is a morpho-syntactic composition of negation and OR (Sauerland 2000; Penka 2011; Zeijlstra 2011). For our purposes here we set aside what is involved in cases of apparent lexicalizations of NOR and why there are no apparent lexicalizations of NAND (Horn 1972), and focus on why no connective other than AND and OR is lexicalized as simplex.

	AND	OR	NOR	NAND	XOR	IFF	NOTL	NOTR	ONLYL	ONLYR	→	←
$p \wedge q$	1	1	0	0	0	1	0	0	0	0	1	1
$p \wedge \neg q$	0	1	0	1	1	0	0	1	1	0	0	1
$q \wedge \neg p$	0	1	0	1	1	0	1	0	0	1	1	0
$\neg p \wedge \neg q$	0	0	1	1	0	1	1	1	0	0	1	1

Table 1: All binary logical connectives which are 'meaningful' (excluding trivial and redundant connectives)

Since Horn (1972), the simplicity of AND and OR has usually been taken for granted, most commonly by assuming that 'positive' connectives, like AND and OR, are inherently less

*For helpful discussions and comments we thank Amir Anvari, Adi Behar Medrano, Milica Denić, Émile Enguehard, Danny Fox, Aron Hirsch, Ezer Rasin and Benjamin Spector, as well as anonymous reviewers and audiences at ILLC, Institut Jean Nicod, University of Leipzig, Tel Aviv University, ZAS Berlin, workshop on Alternatives, expectations and domain-widening at Bar-Ilan University, and IATL 37.

[1] We assume that lexicalization of the non-meaningful connectives in the following table can be ruled out on independent grounds, given arguments that there are general bans on redundancy and triviality (see Gajewski 2002; Sauerland 2004, a.o.).

	L	R	TAU	CONT
$p \wedge q$	1	1	1	0
$p \wedge \neg q$	1	0	1	0
$q \wedge \neg p$	0	1	1	0
$\neg p \wedge \neg q$	0	0	1	0

complex than 'negative' ones, like NAND (see Katzir & Singh 2013, a.o.); we will call this assumption Positivity. As Enguehard & Spector (2021) and Incurvati & Sbardolini (2020) point out, however, there is no compelling independent evidence for Positivity, and it assumes what needs to be explained. Enguehard & Spector propose an account which does not rely on Positivity, but simulations by Züfle & Katzir (2022) suggest that their account can only provide a general theory of the typological pattern if Positivity is assumed after all. Incurvati & Sbardolini also aim to account for the typological pattern without assuming Positivity, but they rely on specific assumptions about the dynamics of conversation which as far as we can see are not independently motivated.[2] More importantly, both accounts wrongly predict NOR to be lexicalizable as simplex.

Our goal is to derive Positivity by building on a property of AND and OR which falls out of an independently needed theory of alternatives and communication. The theory we assume is one where grammar incorporates an exhaustivity mechanism (with no access to real world probabilities), whereas pragmatics is governed by general rational reasoning (which takes real world probabilities into account). For recent evidence in favor of this view see especially Franke & Bergen (2020); Cremers et al. (2022). In §2 we will identify a property that distinguishes AND and OR from all other meaningful connectives based on the notion of Innocent Exclusion from Fox (2007), and in §3 we will use a simple model of communication with access to prior probabilities in order to illustrate the ramifications that this property has for communication. In §4 we will then argue that, as a result, languages with AND and OR (and no other connective) have a communicative advantage over languages with other meaningful connectives, using the notion of Stability from Bar-Lev & Katzir (2022).

2 Exhaustivity

Following Fox (2007), we assume a covert exhaustivity operator $\mathcal{E}xh$ which negates alternatives that are innocently excludable (IE), as defined in (1) (both $\mathcal{E}xh$ and Innocent Exclusion have roots in Groenendijk & Stokhof 1984; see Spector 2016).

(1) a. $[\![\mathcal{E}xh]\!](C)(p)(w) = 1$ *iff* $p(w) = 1 \wedge \forall q \in IE(C,p)[q(w) = 0]$
b. $IE(C,p) = \bigcap\{C' : C'$ is a maximal subset of C s.t. $p \wedge \bigwedge\{\neg q : q \in C'\} \not\Rightarrow_L \bot\}$

Innocent Exclusion aims to avoid arbitrary choices of exclusion. We can describe what (1) does given a set of alternatives C and a prejacent p using the following procedure (see also Bar-Lev & Fox 2017):

(2) **Innocent Exclusion procedure**:
a. Take all maximal sets of alternatives in C that can be assigned false consistently with the prejacent p.
b. Only exclude those alternatives that are members in all such sets—the Innocently Excludable alternatives.

We observe that AND and OR have a property which sets them apart from all other connectives in table 1 and which has to do with Innocent Exclusion: When a sentence of the form P CON

[2]Specifically, they assume a dynamic system where assertion and rejection have a fundamentally different status, and as such their account relies on non-trivial assumptions about the building blocks of the dynamics of conversation (we thank Benjamin Spector for a discussion of this issue). Moreover, it is crucial for their account that rejecting a conjunction should end up equivalent to rejecting a disjunction; we however do not know of evidence that could support this assumption.

Q has a logically independent alternative (LIA) R — that is, an alternative that stands in no entailment relations to either P or Q — the alternative R is IE if CON is AND or OR, but non-IE if CON is one of the other 10 connectives in table 1. We assume that when R is salient P CON Q has the set of alternatives in (3), which is what is predicted by the theory of structural alternatives by Katzir (2007).[3]

(3) $Alt(\text{P CON Q})^{+} = \{x \text{ CON } y : x, y \in \{\text{P,Q,R}\}\} \cup \{\text{P,Q,R}\}$

For instance, R is non-IE given P NAND Q: (P NAND Q)$\wedge\neg$(P NAND R) is consistent, which means there is a maximal set of alternatives whose exclusion is consistent with P NAND Q that contains P NAND R; this set cannot contain R, because (P NAND Q)$\wedge\neg$(P NAND R)$\wedge\neg$R is inconsistent. As a result, R is non-IE. The reader is invited to verify that similar considerations make R non-IE given any sentence of the form P CON Q for any meaningful CON other than AND or OR. Note that no addition of alternatives to $Alt(\text{P CON Q})^{+}$ can make R IE if it isn't, so exclusion of LIAs is only guaranteed in languages without any connective other than AND and OR. In other words, among all 4095 ($= 2^{12} - 1$) possible inventories of meaningful connectives, there are only 3 where LIAs are IE for every sentence of the form P CON Q: {AND}, {OR}, and {AND, OR} (see Uegaki 2022 for evidence that all of them are attested). Excludability of LIAs then distinguishes AND and OR from all other meaningful connectives:

(4) **Excludability of LIAs**: AND and OR are the only meaningful connectives which enable the innocent exclusion of LIAs.

In what follows we will argue that the excludability of LIAs has implications for communication which can explain Positivity.

3 Anti-exhaustivity

Why should the excludability of LIAs matter? We claim that it matters once we consider the implications of non-excludability of LIAs within a theory of communication based on Iterated Rationality Models (IRMs; Franke 2009; Frank & Goodman 2012; Bergen et al. 2016), due to what Cremers et al. (2022) call Anti-exhaustivity. While IRMs have often been used to derive scalar implicatures (SIs), recent work has argued that SIs are grammatically derived, and IRMs should serve as a disambiguation mechanism (see Champollion et al. 2019; Fox & Katzir 2021; Asherov et al. 2022; Cremers et al. 2022). Specifically, Cremers et al. show that, in the absence of grammatical strengthening, IRMs wrongly predict a sentence like P to sometimes end up meaning $p \wedge \neg q$ (exhaustive meaning) and sometimes end up meaning $p \wedge q$ (anti-exhaustive meaning), depending on the prior probabilities. If $\mathcal{E}xh$ applies, however, this wrong prediction disappears, because $\mathcal{E}xh(\text{P})$ excludes the LIA Q, which results in a meaning that is compatible with $p \wedge \neg q$ but incompatible with $p \wedge q$.

Let us illustrate how this problem is relevant for our discussion of connectives and LIAs by considering (a subset of) the set of alternatives in (3) when the connective CON is AND and looking at the predictions of a simple IRM (taken from Bar-Lev & Katzir 2022; this particular choice is not crucial, since as Cremers et al. 2022 pointed out, the issue of Anti-exhaustivity in the absence of grammatical strengthening arises with all IRMs they considered).

The rational hearer we assume is defined in (5-c). Given a certain message, this hearer applies Bayes' Rule to infer the likeliest state(s) given the message, on the assumption that the

[3]For simplicity we ignore replacements of the connective with other connectives; nothing hinges on this though.

message was sent by a rational speaker; if there is exactly one such state, the rational hearer chooses it, i.e., assigns it probability 1 (and if there are several likeliest states the assigned probability ends up evenly divided between them). The rational speaker we assume is defined in (5-b). If one message has the best chances of leading a naive hearer to choose the state that the rational speaker wants to convey, then they assign probability 1 to that message (if there are several best messages the assigned probability ends up evenly divided between them). The naive hearer, in turn, attributes a probability to a state upon hearing a message based on the prior probability of that state (given by the function PR) and whether the message is true in that state (that is, if the message is false in that state the assigned probability is 0).

(5) a. A naive hearer NH: $NH(s|m) \propto PR(s) \cdot [\![\mathrm{m}]\!](s)$

b. A rational speaker RS:

$$RS(m|s) \propto \begin{cases} 1 & \text{if } m \in \underset{m'}{\operatorname{argmax}} NH(s|m') \text{ and } \underset{m'}{\max} NH(s|m') \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

c. A rational hearer RH:

$$RH(s|m) \propto \begin{cases} 1 & \text{if } s \in \underset{s'}{\operatorname{argmax}}(RS(m|s') \cdot PR(s)) \text{ and } \underset{s'}{\max} RS(m|s') \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

The set of messages given by (3) when CON is AND is as follows:[4]

(6) {P AND Q, P AND R, Q AND R, P, Q, R}

To keep the current illustration simple, we will ignore most of the messages in (6) and pretend that the possible messages are just P AND Q and its LIA R (the reader may verify that Anti-exhaustivity will arise even when the full set of messages in (6) is considered). Imagine now a rational speaker who wants to convey the state $p \wedge q \wedge r$, assuming that the prior probabilities of the states (given by the partition of logical space induced by (6)) to be as in Table 2. Naive hearer behavior in this case is predicted to be as in Table 3, and rational speaker behavior is predicted to be as in table 4.

	$p \wedge q \wedge r$	$\neg p \wedge q \wedge r$	$p \wedge \neg q \wedge r$	$\neg p \wedge \neg q \wedge r$	$p \wedge q \wedge \neg r$	$\neg p \wedge q \wedge \neg r$	$p \wedge \neg q \wedge \neg r$	$\neg p \wedge \neg q \wedge \neg r$
Priors	0.2	0.1	0.1	0.1	0.05	0.05	0.05	0.35

Table 2: Prior probabilities of states

NH	$p \wedge q \wedge r$	$\neg p \wedge q \wedge r$	$p \wedge \neg q \wedge r$	$\neg p \wedge \neg q \wedge r$	$p \wedge q \wedge \neg r$	$\neg p \wedge q \wedge \neg r$	$p \wedge \neg q \wedge \neg r$	$\neg p \wedge \neg q \wedge \neg r$
P AND Q	0.8	0	0	0	0.2	0	0	0
R	0.4	0.2	0.2	0.2	0	0	0	0

Table 3: Naive hearer behavior

[4] We only aim to provide here a demonstration of the problem of Anti-exhaustivity, so we ignore the possibility of having more complex alternatives like P AND Q AND R and the effects of message costs. As Cremers et al. (2022) show, anti-exhaustive readings are predicted even with more complex alternatives and sophisticated assumptions about the role of message costs in the model. This however depends on the assumption that message costs should play a role in the model, which we cannot argue for here and have to leave for another occasion.

RS	$p \wedge q$ $\wedge r$	$\neg p \wedge q$ $\wedge r$	$p \wedge \neg q$ $\wedge r$	$\neg p \wedge \neg q$ $\wedge r$	$p \wedge q$ $\wedge \neg r$	$\neg p \wedge q$ $\wedge \neg r$	$p \wedge \neg q$ $\wedge \neg r$	$\neg p \wedge \neg q$ $\wedge \neg r$
P AND Q	1	0	0	0	1	0	0	0
R	0	1	1	1	0	0	0	0

Table 4: Rational speaker behavior

Given a message, the rational hearer takes all states where this message would be an optimal one for the rational speaker and chooses the one which has the highest prior probability (if there is one). The rational hearer's inferences are summarized in table 5.

RH	$p \wedge q$ $\wedge r$	$\neg p \wedge q$ $\wedge r$	$p \wedge \neg q$ $\wedge r$	$\neg p \wedge \neg q$ $\wedge r$	$p \wedge q$ $\wedge \neg r$	$\neg p \wedge q$ $\wedge \neg r$	$p \wedge \neg q$ $\wedge \neg r$	$\neg p \wedge \neg q$ $\wedge \neg r$
P AND Q	1	0	0	0	0	0	0	0
R	0	1/3	1/3	1/3	0	0	0	0

Table 5: Rational hearer behavior

In this case then, the optimal choice for a rational speaker who wants to convey the state $p \wedge q \wedge r$ is P AND Q, and the rational hearer who gets this message will choose $p \wedge q \wedge r$. The model then predicts that an anti-exhaustive interpretation of P AND Q would be possible, something which never actually happens (as Cremers et al. 2022 argue for slightly different cases). Note the importance of priors here: Switching between the prior probabilities of the states $p \wedge q \wedge r$ and $p \wedge q \wedge \neg r$ in table 2 would lead a rational hearer to choose the state $p \wedge q \wedge \neg r$ when they hear P AND Q; in other words, they will have an exhaustive interpretation. Of course, if P AND Q were exhaustified, an anti-exhaustive interpretation would be impossible no matter what the priors look like, because $\mathcal{E}xh$(P AND Q) is only compatible with $p \wedge q \wedge \neg r$ given that the logically independent alternative R is IE: exhaustivity would block anti-exhaustivity.

Recall now that we have shown in §2 that $\mathcal{E}xh$ does not always allow for exclusion of LIAs for sentences with connectives: It only does so if the connective at stake is AND or OR. P NOR Q, for instance, cannot innocently exclude R when exhaustified, and as a result would sometimes end up meaning $\neg p \wedge \neg q \wedge \neg r$ (exhaustive meaning) and sometimes end up meaning $\neg p \wedge \neg q \wedge r$ (anti-exhaustive meaning), depending on the prior probabilities, even if $\mathcal{E}xh$ applies.[5] Similar results hold of any meaningful connective other than AND and OR; for reasons of space we cannot show this here in detail. Under the assumption that IRMs are a good model of communication when taken together with a grammar that incorporates $\mathcal{E}xh$ (a view defended by Franke & Bergen 2020; Cremers et al. 2022), then, we can distinguish between AND/OR and all the other 10 connectives in table 1 as follows:

(7) **Anti-exhaustivity with connectives**: For a meaningful CON, $\mathcal{E}xh$(P CON Q) can have both exhaustive and anti-exhaustive meanings (depending on prior probabilities) *iff* CON $\notin$ {AND, OR}.

4 Stability

Using Bar-Lev & Katzir's (2022) terms (extended here to apply to hearer's choices rather than speaker's choices), a sentence whose interpretation depends on prior probabilities is *unstable*; we assume as in Bar-Lev & Katzir (2022) that instability is a bad property for communication:

[5] To keep things simple we pretend that $\mathcal{E}xh$ always applies and excludes R when it is IE. Our view is however in line with the natural assumption that exclusion of LIAs is not derived when they are irrelevant, in which case Anti-exhaustivity cannot arise (Cremers et al.).

It means that a speaker and a hearer may miscommunicate if they do not share the same assumptions about prior probabilities. We hypothesize that this, in turn, can affect language change: Unstable languages will be filtered out when languages evolve because they are not efficient for communication.

Due to Anti-exhaustivity, languages with connectives other than AND and OR are unstable: Expectations determine whether the interpretation is exhaustive or anti-exhaustive. Based on certain expectations, a speaker may utter P NAND Q, expecting the hearer to infer ¬R (exhaustive interpretation). But a hearer with different prior beliefs may infer R instead (anti-exhaustive interpretation). Languages with AND, OR, or both (but no other meaningful connective) are stable because in such languages exhaustification blocks anti-exhaustive readings: A hearer will never infer R from P AND/OR Q. The difference between AND and OR in terms of excludability of LIAs, which as we have seen has consequences for anti-exhaustivity, leads to the following generalization:

(8) **Stability**: A language (an inventory of connectives) is stable *iff* it contains no meaningful connectives other than AND and OR.

We propose then that AND and OR are the only meaningful connectives for which simplex lexicalizations are attested because they have a communicative advantage, being the only ones that can avoid anti-exhaustivity and consequently instability.

Table 6 summarizes our proposal. AND and OR enable the exclusion of LIAs through exhaustivity, which blocks anti-exhaustivity and leads to stability. Other meaningful connectives do not enable the exclusion of LIAs through exhaustivity, and as a result they give rise to anti-exhaustivity, and any inventories that include such connectives are unstable.

	Exclusion of LIAs		Anti-exhaustivity		Stability
AND and OR	succeeds	⇒	blocked	⇒	✓
Other CONs	fails	⇒	possible	⇒	✘

Table 6: Summary of our proposal

5 Comparison with Bar-Lev & Katzir (2022)

Our view, which is based on the interaction of grammatical strengthening with LIAs, avoids several shortcomings of our proposal in Bar-Lev & Katzir (2022). First, in Bar-Lev & Katzir (2022) we assumed that IRMs apply with no underlying grammatical strengthening; as a result, that system leads to anti-exhaustive readings when LIAs are considered even in a language like {AND, OR} (which then ends up as unstable from the speaker's perspective). In other words, by not assuming grammatical strengthening, the system in Bar-Lev & Katzir (2022) predicts anti-exhaustive readings to arise in English and many other languages, something which, as Cremers et al. (2022) argue, never actually happens.

Second, in Bar-Lev & Katzir (2022) non-commutative connectives are stipulated to be non-lexicalizable (as in all other attempts at explaining the typology we are aware of, see Uegaki 2022 for discussion). $\mathcal{E}xh$, which is needed anyway, makes it possible to dispense with such stipulations: All non-commutative connectives in table 1 lead to Anti-exhaustivity and instability.

Finally, in Bar-Lev & Katzir (2022) we stipulated that every language has negation, as that account is contingent on considering stability in both positive and negative sentences. Our current view singles out languages with AND and OR even without this stipulation.

References

Asherov, Daniel, Danny Fox & Roni Katzir. 2022. The only-implicature generalization and its relevance for theories of pragmatics. Ms., MIT and TAU. https://ling.auf.net/lingbuzz/006837.

Bar-Lev, Moshe E. & Danny Fox. 2017. Universal free choice and innocent inclusion. In Dan Burgdorf, Jacob Collard, Sireemas Maspong & Brynhildur Stefánsdóttir (eds.), Semantics and linguistic theory (SALT), vol. 27, 95–115.

Bar-Lev, Moshe E. & Roni Katzir. 2022. Communicative stability and the typology of logical operators. To appear in Linguistic Inquiry.

Bergen, Leon, Roger Levy & Noah Goodman. 2016. Pragmatic reasoning through semantic inference. Semantics and Pragmatics 9.

Champollion, Lucas, Anna Alsop & Ioana Grosu. 2019. Free choice disjunction as a rational speech act. In Semantics and linguistic theory (SALT), vol. 29, 238–257.

Cremers, Alexandre, Ethan G. Wilcox & Benjamin Spector. 2022. Exhaustivity and anti-exhaustivity in the RSA framework: Testing the effect of prior beliefs. doi:10.48550/ARXIV.2202.07023.

Enguehard, Émile & Benjamin Spector. 2021. Explaining gaps in the logical lexicon of natural languages: A decision-theoretic perspective on the square of aristotle. Semantics and Pragmatics 14. 5.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In Uli Sauerland & Penka Stateva (eds.), Presupposition and implicature in compositional semantics, 71–120. London: Palgrave Macmillan UK. doi:10.1057/9780230210752_4.

Fox, Danny & Roni Katzir. 2021. Notes on iterated rationality models of scalar implicatures. Journal of Semantics doi:10.1093/jos/ffab015.

Frank, Michael C & Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. Science 336(6084). 998–998.

Franke, Michael. 2009. Signal to act: Game theory in pragmatics: Universiteit van Amsterdam dissertation.

Franke, Michael & Leon Bergen. 2020. Theory-driven statistical modeling for semantics and pragmatics: A case study on grammatically generated implicature readings. Language 96(2). e77–e96.

Gajewski, Jon. 2002. L-analyticity and natural language. Manuscript, MIT .

Groenendijk, Jeroen Antonius Gerardus & Martin Johan Bastiaan Stokhof. 1984. Studies on the semantics of questions and the pragmatics of answers: Univ. Amsterdam dissertation.

Horn, Laurence Robert. 1972. On the semantic properties of logical operators in english: University of California, Los Angeles dissertation.

Incurvati, Luca & Giorgio Sbardolini. 2020. Update rules and semantic universals.

Katzir, Roni. 2007. Structurally-defined alternatives. Linguistics and Philosophy 30(6). 669–690. doi:10.1007/s10988-008-9029-y.

Katzir, Roni & Raj Singh. 2013. Constraints on the lexicalization of logical operators. Linguistics and Philosophy 36(1). 1–29.

Penka, Doris. 2011. Negative indefinites 32. Oxford University Press.

Sauerland, Uli. 2000. No 'no': on the crosslinguistic absence of a determiner 'no'. In Hidekazu Suzuki (ed.), Proceedings of the Tsukuba Workshop on Determiners and Quantification, 415–444. Tsukuba: Tsukuba University.

Sauerland, Uli. 2004. Scalar implicatures in complex sentences. Linguistics and philosophy 27(3). 367–391. doi:10.1023/B:LING.0000023378.71748.db.

Spector, Benjamin. 2016. Comparing exhaustivity operators. Semantics and Pragmatics 9(11). 1–33. doi:10.3765/sp.9.11.

Uegaki, Wataru. 2022. The Informativeness/Complexity Trade-Off in the Domain of Boolean Connectives. Linguistic Inquiry 1–39. doi:10.1162/ling_a_00461.

Zeijlstra, Hedde. 2011. On the syntactically complex status of negative indefinites. The Journal of Comparative Germanic Linguistics 14(2). 111–138.

Züfle, Maike & Roni Katzir. 2022. Reasoning about stored representations in semantics using the typology of lexicalized quantifiers. In Proceedings of Sinn und Bedeutung 26, .

Expressives do shift, but not on their own

Corien Bary[1], Leopold Hess[2], and Bob van Tiel[3]

[1] Radboud University Nijmegen, Nijmegen, The Netherlands
corien.bary@ru.nl
[2] Jagiellonian University, Kraków, Poland
leopold.hess@uj.edu.pl
[3] Radboud University Nijmegen, Nijmegen, The Netherlands
bob.vantiel@ru.nl

1 Introduction

This paper presents the results of a series of studies concerning the availability of "shifted", i.e. non-speaker-oriented interpretations of expressives. Potts (2005) claimed that the content of expressives (words directly conveying a strong, typically negative, attitude, e.g., 'the bastard') is always speaker-oriented, unless the expressive is part of overt quotation.

This claim was soon falsified by examples like (1) (Amaral et al., 2007):

(1) [Monty's father:] Well, in fact Monty said to me this very morning that he hates to mow the friggin' lawn.

A non-speaker-oriented reading of 'friggin" in (1) is at least possible, if not preferred. To see if such readings are restricted to speech and attitude reports, Harris and Potts (2009) conducted an experiment with vignettes like (2):

(2) My friend Mike said that his housemate threw a horrible party last weekend. The cretin always invites a lot of people.

Participants were asked to indicate if they attributed the content of the expressive ('the cretin') to the speaker or to the subject of the first sentence (i.e., Mike in the example).

Harris and Potts' results showed that non-speaker-oriented readings were systematically available: as for (2), almost 30% selected Mike as the one who thought that the housemate is a cretin (see also Kaiser, 2015). This led Harris and Potts to propose a pragmatic mechanism to account for the apparent possibility of "shifted" expressives outside of direct or indirect reports.

This pragmatic account relies on the assumption that, in examples such as (2), the expressive should not be construed as covertly embedded in a report, even when its content is attributed to the subject, and not the speaker. Consequently, the pragmatic account assumes that the attribution of the content of the expressive is to a large degree independent from the attribution of the content of the clause in which it occurs. This goes against configurational approaches that assume semantic binding by attitude predicates (Schlenker, 2007; Sauerland, 2007).

Hess (2018) further develops the pragmatic account, arguing that the content of an expressive is attributed as commitment *de lingua* (a concept introduced in Harris 2016: a commitment to the appropriateness of a certain expression in a given context), which is independent of assertoric commitments (i.e. ones concerning the main propositional content of a clause).

In what follows, we employ the terminology of commitment attributions, which are to be distinguished from attributions of literal utterances. Thus, (3), with a direct quotation, attributes the exact words to John, while (4) only attributes the commitment to the proposition that Tories will loose.

(3) John said "Those bloody Tories will loose the next election".

(4) John said that the Tories will loose the next election.

(5) A CPJ report on Venezuela tells us how problems have 'escalated' in Venezuela under Chavez, i.e. the physical attacks against journalists under previous presidents have 'escalated' to Chavez calling the opposition, which includes the media, names. This is very, very serious, but I don't think another coup attempt is called for until Chavez resorts to dramatic irony or sarcasm. But if **that vicious bastard** uses litotes, then there's no other rational choice than an immediate invasion.

Similarly, (5) (from a blog post cited in Potts 2007) attributes a commitment *de lingua* to the authors of the CPJ report without suggesting that they actually used these specific words.

We hypothesized that the assumption of the pragmatic account is wrong and that subject-oriented readings of expressives are typically accompanied by the attribution of the content of the whole clause to the subject, as in a covert report (continuing the report from the previous sentence), i.e., that commitment *de lingua* attribution is not independent of assertoric commitment attribution. This hypothesis is inspired by prior research on appositives—another typically speaker-oriented category—by Koev (2014), who found that shifted appositives are usually understood as uttered in a secondary (i.e., reported) speech context. We tested our hypothesis in a series of experimental studies.

2 Experimental studies

We conducted three experiments to investigate whether subject-oriented (i.e., non-speaker-oriented) readings of expressives are typically accompanied by the attribution of the commitment to the content of the whole sentence to the subject.

Exp. 1 replicated Harris and Potts' experiment with some changes. Participants read two-sentence stories such as (2) and had to indicate, on a five-point scale, to what extent they attributed the expressive to the speaker or to the subject of the first sentence.

Exp. 2 used the same materials as Exp. 1, but instead asked participants to indicate, again on a five-point scale, to what extent they attributed the content of the entire second sentence to the speaker or to the subject of the first sentence. If subject-oriented readings of expressives indeed involve covert speech reports, we expect that stories in which the expressive is likely to receive a subject-oriented interpretation are also stories in which the content of the entire second sentence is likely to be attributed to the subject; in other words, we expect a robust correlation between the results of Exps. 1 and 2.

Exp. 3 essentially replicates Exps. 1 and 2 using a within-participants design. Participants in the experiment read the same short stories used in the previous experiments. However, in this experiment, each story was followed by two questions, inquiring (i) to what extent the expressive was attributed to the speaker or the subject (i.e., the question asked in Exp. 1), and (ii) to what extent the content of the entire second clause was attributed to the speaker or the subject (i.e., the question asked in Exp. 2). We predict a substantial correlation between participants' answers to (i) and (ii), i.e., we expect that subject-oriented readings of the expressive are typically accompanied by subject-oriented readings of the entire second sentence.

In the next sections, we describe the experiments in more detail, and discuss the results. All experiments were hosted on the PCIbex Farm (`farm.ibex.net`) (Zehr and Schwarz, 2018).

2.1 Experiment 1: Expressive attribution

2.1.1 Participants

50 participants were recruited on Prolific. Their mean age was 37 (range: 20–73). 28 participants identified as female; 22 as male. One participant was removed from the analysis because they indicated their native language was not English.

2.1.2 Materials and procedure

The materials consisted of 30 two-sentence stories, such as (2). In each story, the first sentence was a speech report describing a negative event (e.g., 'My friend Mike said that his housemate threw a horrible party'). The second sentence always contained an expressive referring to the person being singled out in the speech report (e.g., 'The cretin always invites a lot of people', where 'the cretin' refers to Mike's housemate).

Half of the stories were borrowed from Harris and Potts (2009, Exp. 2); the remaining ones were constructed by us, using Harris and Potts' stories as a template.

Each story was followed by a question. 20 stories (including all of the ones that were borrowed from Harris and Potts, 2009, Exp. 2) were followed by a target question, which asked whether participants attributed the expressive to the speaker or the subject of the first sentence. For example, (2) was paired with the following question:

(6) Who is calling Mike's housemate a cretin?

To answer this question, participants could mark a value on a five-point scale, where 1 indicated a clearly speaker-oriented interpretation ('Clearly me'), and 5 indicated a clearly subject-oriented interpretation ('Clearly Mike'). Fig. 1 shows an example item.

My secretary Cheryl said that her husband made a complete mess of the basement renovation. The oaf spent a lot of money buying equipment.

Who is calling Cheryl's husband an oaf?

Clearly me ○ ○ ○ ○ ○ Clearly Cheryl

Figure 1: Example trial from **Exp. 1**

The remaining 10 stories were followed by control questions that had a clear-cut answer. For 5 stories, the correct answer was the speaker; for 5 stories, the correct answer was the subject of the first sentence. The response options were the same as for target questions. The order of presentation was randomised for each participant.

Exp. 1 differed from Harris and Potts (2009, Exp. 2) in a number of important respects. First, Harris and Potts varied whether or not the expressive in the second sentence was congruent with the information expressed by the subject in the first sentence. For example, in (2), they varied whether Mike's housemate was said to have thrown a horrible party (congruent) or a fantastic party (incongruent). In our stories, the expressive was always congruent, i.e., the first sentence always set up a negative context in line with the negative connotation of the expressive. Second, Harris and Potts varied whether or not the first sentence contained an intensifier (e.g., 'really', 'totally'). We only used stories without intensifiers. Third, Harris

and Potts asked participants *whose view* it is that, e.g., Mike's housemate is a cretin, whereas we ask *who is calling* Mike's housemate a cretin. This was intended to target attribution of commitment *de lingua*, as opposed to a "whose view" question asking only about the attitude expressed, or a "who called" question asking about specific words used.

2.1.3 Results

Results for 10 trials had to be removed because of a technical error. Five participants were removed from the analysis because they made mistakes on more than 20% of the control items. The error rate of the remaining participants was 3%.

The mean rating for target items was 2.95, which is close to the middle of the scale. Speaker-oriented responses (i.e., 1 or 2) were given on 43% of the trials; subject-oriented responses (i.e., 4 or 5) on 38% of the trials. The middle value was marked on 19% of the trials. Mean ratings were relatively stable across target items, ranging from 2.52 to 3.20.

These findings show that, although there was a slight preference for expressives to receive a speaker-oriented interpretation, subject-oriented readings were readily available. In Exp. 2, we investigate whether the availability of subject-oriented readings is correlated with the probability with which the second sentence as a whole is attributed to the subject.

2.2 Experiment 2: Sentence attribution

2.2.1 Participants

50 participants were recruited on Prolific. Their mean age was 34 (range: 19–58). 20 participants identified as female, 23 as male, 7 had a different gender identity. Two participants were removed from the analysis because they indicated their native language was not English.

2.2.2 Materials and procedure

The materials were the same as for Exp. 1. However, in this experiment, the target question asked whether participants attributed the content of the second sentence to the speaker or the subject of the first sentence. For example, (2) was paired with the following question:

(7) Who is claiming that Mike's housemate always invites a lot of people?

As in Exp. 1, participants answered by marking a value on a five-point scale, where 1 indicated a clearly speaker-oriented interpretation ('Clearly me'), and 5 indicated a clearly subject-oriented interpretation ('Clearly Mike').

The order of presentation was randomised for each participant.

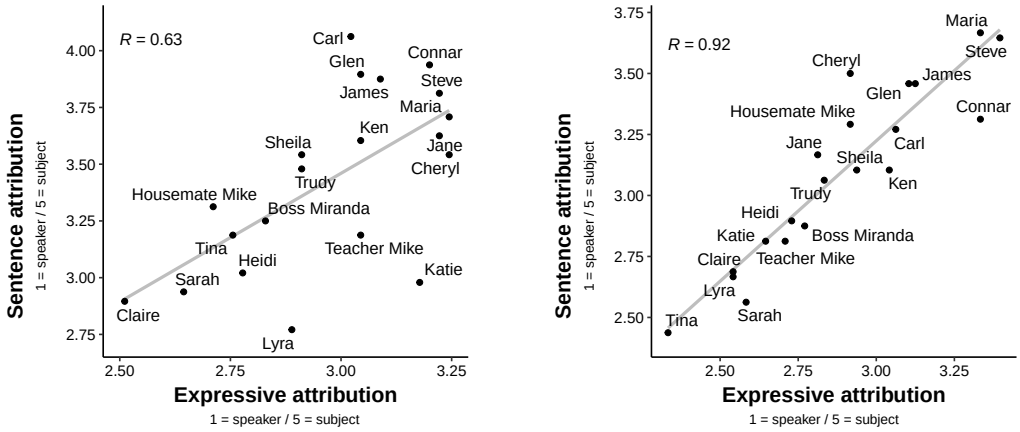
2.2.3 Results

Two participants were removed from the analysis because they made mistakes on more than 20% of the control items. The error rate of the remaining participants was 3%.

The mean rating for target items was 3.44, which indicates a slight preference for attributing the content of the second sentence to the subject rather than the speaker. Such subject-oriented readings suggest that participants readily construed the second sentence as a covert speech report continuing from the previous sentence.

Fig. 2a plots, for each target item, the average ratings in Exp. 1 and Exp. 2. We observed a robust correlation between these ratings ($r(18) = .66$, $p = .002$). Hence, items where the expressive often received a subject-oriented interpretation tended to be items where the content

of the entire second sentence was attributed to the subject, thus supporting the idea that subject-oriented readings of expressives involve covert speech reports.



(a) **Exps. 1 & 2** (b) **Exp. 3**

Figure 2: Scatterplots showing, for each story, the average rating for the expressive (1 = speaker-oriented, 5 = subject-oriented) against the average rating for the whole sentence (1 = speaker-oriented, 5 = subject-oriented).

Exp. 3 seeks to provide more direct evidence by asking the same participants to indicate both the source of the expressive and the source of the second sentence, i.e., by combining Exps. 1 and 2 in a within-participants design.

2.3 Experiment 3: Expressive and sentence content attribution

2.3.1 Participants

50 participants were recruited on Prolific. Their mean age was 37 (range: 21–72). 24 participants identified as female, 20 as male, 2 had a different gender identity. One participant was removed from the analysis because they indicated their native language was not English.

2.3.2 Materials and procedure

The materials were the same as for Exps. 1 and 2. However, in this experiment, each story was followed by two questions rather than one. For target items, the first question asked whether participants attributed the expressive to the speaker or to the subject of the first sentence (i.e., the target question from Exp. 1); the second question asked whether participants attributed the content of the second sentence to the speaker or to the subject of the first sentence (i.e., the target question from Exp. 2). For control items, one question had the speaker as correct answer; the other question had the subject of the first sentence as correct answer.

The order of presentation of the two questions was randomised on each trial. The order of presentation of the stories was randomised for each participant.

2.3.3 Results

For four participants, demographic information was not recorded due to a technical error. Data for these participants was still included in the analysis. Two participants were removed from the analysis for making mistakes on more than 20% of the control items. The error rate of the remaining participants was 3%.

The mean ratings were 2.89 for the expressive attribution question and 3.1 for the sentence attribution question. These means are similar to what we observed in Exps. 1 and 2 (i.e., 2.95 and 3.44).

Fig. 2b plots, for each target item, the average ratings for the two questions. These ratings were strongly correlated ($r(18) = .91$, $p < .001$). We also analysed the overall correlation between participants' ratings to the two questions without aggregating over items. Again, we observed a robust correlation ($r(938) = .71$, $p < .001$).

The latter analysis is the more relevant one, since it shows that participants who attributed the expressive to the subject were also likely to attribute the sentence in which it occurred to the subject, which supports the idea that commitments *de lingua* and assertoric commitments tend to be attributed jointly.

3 Discussion

We corroborated the systematic availability of non-speaker-oriented readings of expressives observed by Harris and Potts. At the same time, our experiments confirmed the hypothesis that the assumption of the pragmatic shift account is wrong: non-speaker-oriented interpretations of expressives are typically accompanied by non-speaker-oriented interpretations of the whole clause in which they occur; i.e. commitments *de lingua* tend to be attributed in concert with assertoric commitments. This suggests that subject-oriented readings of expressives—at least in examples like (7)—should be construed as cases where the expressive is covertly embedded in a report, which undermines the motivation for the pragmatic account.

At the same time, these subject-oriented readings of expressives do not lend themselves easily to a quotational analysis such as proposed by Anand (2007). Quotation is normally entirely flexible in scope and should make it possible to attribute a single word to the subject without attributing the content of the whole sentence (e.g. 'Mike didn't know that the 'cretin' wanted to do something nice for him.').

Our results also shed light on the scope and meaning of the "shift together" constraint posited by Anand and Nevins (2004): the reason that multiple perspectival elements in one clause/domain tend to shift to the same perspective may be simply that the whole clause/domain is interpreted as attributed to someone else. This opens up the possibility of reducing many complex and puzzling semantic phenomena to interactions between *de lingua* and assertoric commitments. The precise relations between commitment attribution on the one hand and mechanisms of (covert) quotation or shift-together constraints requires much further theoretical analysis, as well as empirical investigation.

References

Amaral, P., C. Roberts, and E. A. Smith (2007). The logic of conventional implicatures. *Linguistics and Philosophy 30*(6), 707–749.

Anand, P. (2007). Re-expressing judgment. *Theoretical Linguistics 33*(2), 199–208.

Anand, P. and A. Nevins (2004). Shifty operators in changing contexts. In *Proceedings of SALT*, pp. 20–37.

Harris, J. (2016). Commitments by proxy: Perspective management with transparent free relatives. GLOW Workshop on Perspectivization, Göttingen. www.uni-goettingen.de/de/document/download/25b7960a31a7c1fef02b94d6c5c4dccc.pdf/Harris-GLOW2016-TFRs.pdf.

Harris, J. and C. Potts (2009). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy 32*(6), 523–552.

Hess, L. (2018). Perspectival expressives. *Journal of Pragmatics 129*, 13–33.

Kaiser, E. (2015, 12). Perspective-shifting and free indirect discourse: Experimental investigations. *Semantics and Linguistic Theory 25*, 346.

Koev, T. (2014). Two puzzles about appositives: Projection and perspective shift. In U. Etxeberria, A. Fălăuş, A. Irurtzun, and B. Leferman (Eds.), *Proceedings of Sinn und Bedeutung 18*, Bayonne and Vitoria-Gasteiz, pp. 217–234.

Potts, C. (2005). *The logic of conventional implicatures*. Oxford studies in theoretical linguistics. Oxford: Oxford University Press.

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics 33*(2), 165–198.

Sauerland, U. (2007). Beyond unpluggability. *Theoretical Linguistics 33*(2), 231–236.

Schlenker, P. (2007). Expressive presuppositions. *Theoretical Linguistics 33*(2), 237–245.

Zehr, J. and F. Schwarz (2018). PennController for Internet Based Experiments (IBEX).

As complex as they appear: Children's comprehension of conjunctive expressions in Georgian*

Cory Bill[1], Aurore Gonzalez[2], Imke Driemel[3], and Tamar Makharoblidze[4]

[1] Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany
bill@leibniz-zas.de
[2] University of Milano-Biccoca, Milan, Italy.
aurore.gonzalez@unimib.it
[3] Humboldt-Universität zu Berlin, Berlin, Germany
imke.driemel@hu-berlin.de
[4] Ilia State University, Tbilisi, Georgia
tamar.makharoblidze@iliauni.edu.ge

Abstract

Mitrović and Sauerland [2014, 2016] claim that, across languages, DP-conjunction decomposes into [$_{JP}$ [$_{MuP}$ [XP$_1$ ⟲] MU] [$_{J'}$ J [$_{MuP}$ [XP$_2$ ⟲] MU]]]. The J-particle which occurs between conjuncts maps to set intersection, while the MU-particles map to the subset operator. Given that expressions which involve non-pronounced elements should be harder for children to comprehend than expressions which do not have such silent elements [Slobin, 1985, van Hout, 2008, Sauerland and Alexiadou, 2020, Guasti et al., 2022, a.o.], the account by Mitrović and Sauerland [2014, 2016] predicts that sentences where either J or MU are pronounced should be harder to comprehend relative to sentences where both J and MU are realized. We conduct an experiment testing this prediction by investigating children's comprehension of conjunctive expressions in Georgian. Our results constitute a challenge for the account by Mitrović and Sauerland [2014, 2016] and other existing accounts of conjunctive expressions [Brasoveanu and Szabolcsi, 2013, Szabolcsi, 2015, a.o.]

1 Introduction

One of the basic features of human language is the ability to create conjunctive expressions, that is, to build expressions through the joining of two (or more) phrases, as shown in (1).

(1) a. Mary and Susan love cars. DP
b. Mary loves cars and hates scooters. VP
c. Mary loves cars and Susan hates scooters. TP

Mitrović and Sauerland [2014, 2016] claim that languages share the same underlying structure for sentences involving DP-coordination like (1a). That is, across languages, DP-conjunction decomposes into [$_{JP}$ [$_{MuP}$ [XP$_1$ ⟲] MU] [$_{J'}$ J [$_{MuP}$ [XP$_2$ ⟲] MU]]]. The J-particle

*We would like to thank Lucie Petit for creating the experiment pictures, and Maria Teresa Guasti and Lilla Pintér for feedback. Ethical review of the experimental protocol was done by the review board of Ilia State University and the ERC Executive Agency. Consent was obtained from participants. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 856421).
Authorship: CB and AG conceived the experimental question, designed the experiment, conducted analyses, and drafted manuscript. ID conceived the experimental question, designed the experiment, and conducted typological analysis. TM designed the experiment and collected Georgian data. All authors revised and accepted the final version of the paper.

which occurs between conjuncts maps to set intersection, while the MU-particles which attach to each individual conjunct map to the subset operator. Before combining with the MU-particles, each conjunct combines with a type-shifter $\circlearrowleft$ which turns the individual into the characteristic property of that individual.

Languages are proposed to vary with regard to which of these particles they realize. While some languages pronounce J (e.g., English *and*), others pronounce MU (e.g., Japanese *mo mo*). Very few languages realize both J and MU in the same utterance. To our knowledge, the only languages reported so far are Hungarian, South-Macedonian, and Avar [Szabolcsi, 2015, Mitrović and Sauerland, 2016, a.o.].

This work focuses on Georgian, a language that seems to share this feature of having two conjunctive particles at its disposal: the particle *da* and the clitic particle *-c* which also has an additive use [Hewitt, 1995, Chutkerashvili, 2009]. Our investigations reveal that Georgian shows triadic exponence of J and MU, as shown in (2).

(2) *Georgian*

a. *k'ovz-i* **da** *saban-i* J
spoon-NOM J blanket-NOM

'the spoon and the blanket'

b. *k'ovz-i*-**c** *saban-i*-**c** MU
spoon-NOM-MU blanket-NOM-MU

'the spoon and the blanket'

c. *k'ovz-i*-**c** **da** *saban-i*-**c** J-MU
spoon-NOM-MU J blanket-NOM-MU

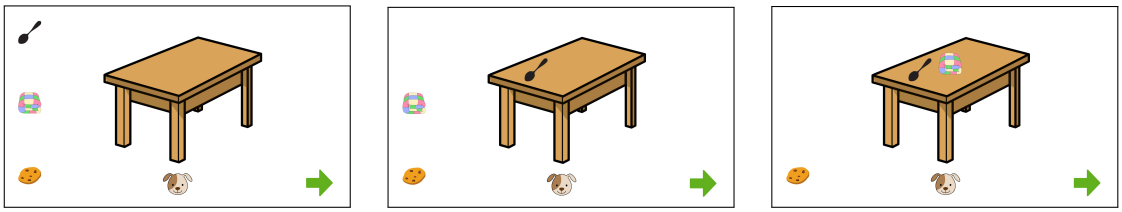
'the spoon and the blanket'

The aim of this work is to test Mitrović and Sauerland's (2014, 2016) account by investigating the comprehension of conjunctive expressions by Georgian-speaking children. Previous work in the acquisition literature indicates a preference from children for expressions that display more one-to-one mapping between form and meaning [Slobin, 1985, van Hout, 2008, Sauerland and Alexiadou, 2020, Guasti et al., 2022, a.o.]. This preference, combined with the account by Mitrović and Sauerland [2014, 2016], predicts that the expressions where both J and MU particles are articulated (i.e., (2c)) should be easier for children to comprehend relative to expressions where one of these particles is silent (i.e., (2a) and (2b)).

The remainder of the paper is organized as follows. Section 2 presents the experiment, the findings of which will be presented in Section 3. Section 4 discusses those findings and concludes.

2 Method

The goal of this experiment was to test the prediction that Georgian-speaking children would display better performance in comprehending conjunctive expressions containing both MU and J particles, relative to conjunctive expressions containing only MU or only J particles. We conducted an experiment with Georgian-speaking adults and children. Note that, before any data was collected, a preregistration of this experiment was published on the Open Science Foundation website and can be found here: https://osf.io/ve9n8.



(a) None (b) One (c) Both

Figure 1: Various starting layouts for the sentences in (3)

2.1 Participants

Thirty-one Georgian-speaking children (3;9-5;10, M = 4;9) and 41 Georgian-speaking adults participated in the experiment. Georgian-speaking children were recruited from daycare centers in Ozurgeti, whereas Georgian-speaking adults were students recruited at Ilia State University.

2.2 Paradigm

We used an act-out task. Each trial began with the participant being presented with one of the starting layouts in Figure 1. Participants then pressed the dog face and were played a pre-recorded version of one of the sentences in (3). Participants could replay the pre-recorded sentence as many times as they wished. After listening to the sentence, participants were instructed that they should change the scene to make the picture match the sentence, if it did not match already. They could do this by moving objects on and off the table.

There were three different *starting layouts* for the objects, which varied with regard to the number of conjuncts that were already satisfied. Specifically: NONE (Fig. 1a), ONE (Fig. 1b), and BOTH (Fig. 1c). We also presented three *sentence-types*: J (3a), MU (3b), and J-MU (3c).

(3) *Georgian*

a. *k'ovz-i* **da** *saban-i* *aris* *magida-ze.* J
spoon-NOM J blanket-NOM is table-on

'The spoon and the blanket are on the table.'

b. *k'ovz-i***-c** *saban-i***-c** *aris* *magida-ze.* MU
spoon-NOM-MU blanket-NOM-MU is table-on

'The spoon and the blanket are on the table.'

c. *k'ovz-i***-c** **da** *saban-i***-c** *aris* *magida-ze.* J-MU
spoon-NOM-MU J blanket-NOM-MU is table-on

'The spoon and the blanket are on the table.'

Crossing the 3 *starting layouts* in Figure 1 and the 3 *sentences-types* in (3) results in 9 conditions. We had 2 items per condition, resulting in a total of 18 items. The items were created as a joint effort between the local researcher based in Georgia and the remote researchers based in Germany and Italy. We included no fillers based on the worry that further items would extend the experimental session beyond children's abilities. Every participant was presented with every item. We measured response *accuracy*; specifically, whether the response was consistent

with an exhaustified interpretation of the sentence (e.g., the spoon and the blanket are on the table, and nothing else is on the table).

2.3 Procedure

Children were tested in a quiet room, at the institution where they were recruited, with an experimenter present. The experiment was displayed using a presentation software on a tablet. A warm-up phase consisting of 4 items, allowing the child to understand the task and get used to the tablet, was followed by the experimental phase consisting of 18 items in randomized order. The experimental sessions were conducted by the local researcher and their affiliates. Adult participants were presented with an on-line version of the experiment, with the instructions presented via text.

3 Results

We conducted two analyses of our data. The first corresponds to the planned analyses described in our preregistration (https://osf.io/ve9n8).

The second analysis is exploratory, and includes an analysis of a new measure variable, *sentence played*, which records how many times the target sentence was played for a given trial.

3.1 Planned analysis

As planned, we investigated the prediction that children's *accuracy* would be better when presented with J-MU sentences compared with MU or J sentences. Figure 2 plots the outcome of this measure variable. Following Barr et al. [2013], we generated a mixed-effects logistic regression analysis with the maximal random effect structure justified by the design. We then reduced this structure in order to achieve convergence by removing the elements with the least variance. This resulted in a model with an intercept for item and an intercept for subject. The fixed effect structure contained *group*, *sentence-type* and their interaction.[1] We then tested for the significance of the fixed effects by comparing this full model to a model that was equivalent except that it was missing one of the fixed effects. We found a significant effect of group ($\chi^2(1) = 12.27$, $p < 0.001$), with adults achieving higher rates of accuracy than children (see Figure 2). However, we did not find a significant effect of *sentence* ($\chi^2(2) = 2.24$, $p = 0.33$) nor the predicted significant interaction of *group* and *sentence* ($\chi^2(2) = 1.95$, $p = 0.38$).[2]

As mentioned in Section 2, participants could replay the pre-recorded sentence as many times as they liked. It is possible that allowing participants to do this generated a type of 'ceiling effect' whereby any difference in the comprehension *accuracy* of the different conjunctive expressions was cancelled.

If this explanation is on the right track, then we might expect the difference between these sentence to rather show up in the *sentence-played* measure; a measure identifying how many times a sentence was played in each trial. Specifically, we might expect that sentences that were more difficult to comprehend were played more times than those that were easier to comprehend. We investigated this possibility further in our exploratory analysis, which we turn to now.

[1] As presented in our preregistration (https://osf.io/ve9n8), we originally also included *sentence layout* in our fixed effect structure, however, we ultimately removed it because; i) with it, the model was too complex to achieve convergence, and ii) we did not make any specific predictions about it.

[2] Note, that we removed one of the items in the NONE.J-MU condition, due to a substantial number of participants expressing confusion about the nature of one of the objects presented in this item.

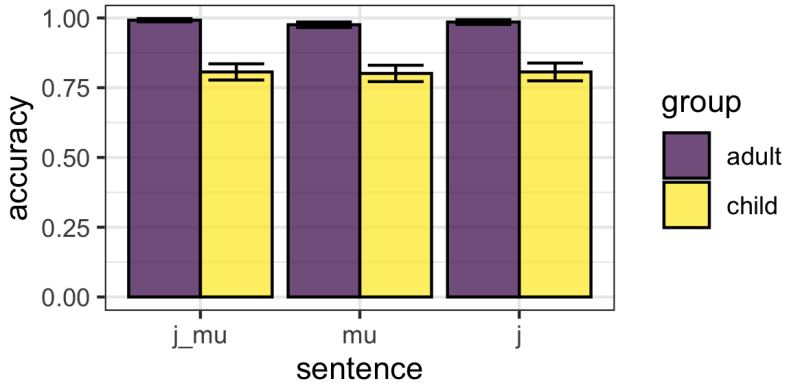


Figure 2: Mean accuracy of end-state. The vertical bars represent the standard error.

3.2 Exploratory analysis

The rates of *sentence played* across sentences and groups is presented in Figure 3.[3] Once again, following Barr et al. [2013], we constructed a linear mixed-effect model with a random intercept for subject. The model also had fixed effects for *group*, *sentence-type* and their interaction. We then determined the statistical significance of the fixed effects via model comparison. We found a significant effect for *group* ($\chi^2(1) = 36.82$, $p < 0.001$), *sentence-type* ($\chi^2(2) = 12.71$, $p < 0.01$), and their interaction ($\chi^2(2) = 20.52$, $p < 0.001$). Figure 3 shows that the main effect of *group* is driven by children playing the sentence more times than adults. Follow-up tests on the *sentence-type* effect indicated that, for children, there was a significant difference between J-MU sentences and MU sentences ($t = 3.68$, $p < 0.001$), as well as between J-MU sentences and J sentences ($t = 5.19$, $p < 0.001$). However, no difference was found between MU sentences and J sentences ($t = 1.67$, $p = 0.22$).[4] No differences were found for adults between the different *sentence-types*.

4 Discussion

Recall the prediction of the account by Mitrović and Sauerland [2014, 2016] we were testing in our study: J and MU sentences should be harder to comprehend for children relative to J-MU sentences. As expected, adults successfully completed the task, providing mainly accurate responses. Turning to the children's results, we found that they showed the same rate of accuracy for all types of sentences, i.e., J, MU and J-MU sentences. As mentioned in Section 3.1, we hypothesize that allowing children to replay the target sentence as many times as they liked may have cancelled any difference in the ease of comprehension of the different types of conjunctive expressions. In other words, we do not take this result as going against the account by Mitrović and Sauerland [2014, 2016]. This interpretation is corroborated by the fact that in the exploratory analysis on the *sentence played* measure, children were found to have played sentences with J-MU expressions more times than sentences with J or MU expressions. One may wonder whether these results relate to the frequency of these expressions in Georgian: children may have been prompted to replay a sentence more times when that sentence is less

[3] We excluded data points that were more than 3 standard deviations from the mean.
[4] P-values were adjusted for multiple comparisons via the Tukey method.

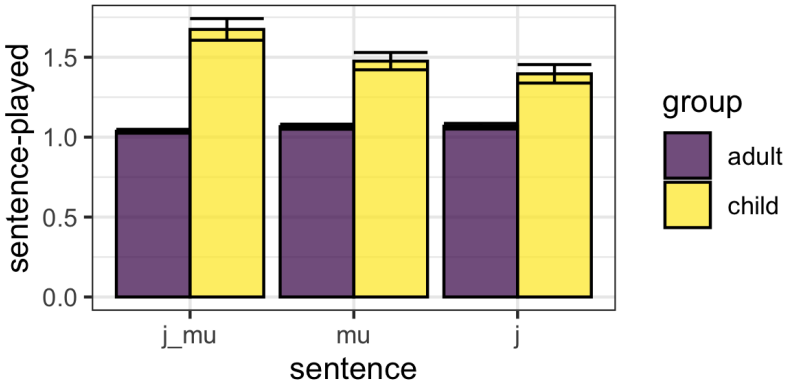


Figure 3: Mean number of times the sentence was played. The vertical bars represent the standard error.

frequently produced in the language generally. This, however, does not seem on the right track given that MU expressions are the less commonly used conjunctive expressions in Georgian.[5] Another potential explanation that can be ruled out on the basis of our results concerns the morphological complexity of these conjunctive expressions. If children were prompted to replay morphologically complex conjunctive expressions more often than the others, they should have played sentences with MU expressions more often than sentences with J expressions, contrary to our results.

Taking the *sentence played* variable as a measure of how easily sentences are understood by children, this study shows that (i) J-MU expressions are more difficult to comprehend than J expressions, (ii) J-MU expressions are more difficult to comprehend than MU expressions. These results constitute a challenge for the account by Mitrović and Sauerland [2014, 2016] which predicted the opposite. In contrast, (i) is in line with accounts in which J expressions and J-MU expressions do not share the same underlying structure. In particular, the underlying representation of J expressions only involves a J particle, unlike that of J-MU expressions which involves both J and MU-particles [Brasoveanu and Szabolcsi, 2013, Szabolcsi, 2015, a.o]. On such accounts, J-MU expressions are underlyingly more complex (i.e., involve more logical operators) than J expressions which would explain why J-MU expressions are harder to comprehend for children than J expressions. An additional complexity about J-MU expressions could come from the presuppositions they come with. Extending the account by Brasoveanu and Szabolcsi [2013] or Szabolcsi [2015] to Georgian would result in analyzing MU particles as (a special kind of) additive particles. As a result, J-MU expressions, unlike J expressions, would come with some additive presuppositions and in turn involve another level of complexity for children. Turning to the result in (ii), this finding constitutes to our knowledge a challenge for all current existing accounts of conjunctive expressions including Brasoveanu and Szabolcsi [2013] and Szabolcsi [2015], which posit a silent J-particle in MU expressions. On such accounts, MU expressions share the same underlying structure as J-MU expressions. Thus, we would expect J-MU expressions where both J and MU are articulated to be easier for children to comprehend than MU expressions where only MU is articulated, contrary to our results. Therefore, our study shows that an alternative account of conjunctive expressions in Georgian is required.

[5] Tamar Makharoblidze, p.c.

References

Dale J. Barr, Rogers Levy, Christoph Scheepers, and Harry J. Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3):255 – 278, 2013. ISSN 0749-596X.

Adrian Brasoveanu and Anna Szabolcsi. Presuppositional Too, Postsuppositional Too. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *The dynamic, inquisitive, and visionary life of* Φ*,* $?\Phi$*, and* $\diamond\Phi$*: a festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman*, pages 55–64. ILLC Publications, Berlin/Boston, 2013.

Anna Chutkerashvili. On -c and ḳi Particles in Georgian. In Peter Bosch, David Gabelaia, and Jérôme Lang, editors, *7th International Tbilisi Symposium on Logic, Language, and Computation*, pages 62–68. Springer, Berlin, 2009.

Maria Teresa Guasti, Artemis Alexiadou, and Uli Sauerland. Undercompression errors as evidence for conceptual primitives. Ms., 2022.

B. G. Hewitt. *Georgian: a structural reference grammar.* John Benjamins, Amsterdam/Philadelphia, 1995.

Moreno Mitrović and Uli Sauerland. Decomposing coordination. In Jyoti Iyer and Leland Kusmer, editors, *Proceedings of NELS 44*, pages 39–52. GLSA, Amherst, 2014.

Moreno Mitrović and Uli Sauerland. Two conjunctions are better than one. *Acta Linguistica Hungarica*, 4:471–494, 2016.

Uli Sauerland and Artemis Alexiadou. Generative grammar: A meaning first approach. *Frontiers in Psychology*, 11:3104, 2020. doi: 10.3389/fpsyg.2020.571295.

Dan I Slobin. Cross-linguistic evidence for the language-making capacity. In Dan I Slobin, editor, *The Cross-linguistic Study of Language Acquisition*, volume 2, pages 1157–1249. Lawrence Erlbaum, Hillsdale, NJ, 1985.

Anna Szabolcsi. What do quantifier particles do? *Linguistics and Philosophy*, 38:159–204, 2015.

Angeliek van Hout. Acquiring perfectivity and telicity in Dutch, Italian and Polish. *Lingua*, 11, 2008.

Antipresuppositions, logophors and shifted indexicality

David Blunier

Université de Genève, Switzerland
david.blunier@unige.ch

Abstract

Some languages can 'shift' indexicals, such as *I* and *you*, and use them anaphorically to refer to participants of the speech event being reported (Schlenker 2003, Deal 2020); other languages use dedicated 'logophoric' pronouns in the same fashion. A fact about logophoric languages is that, in speech reports, the use of a 3rd person form to refer to the author of the report is prohibited, giving rise to a disjointness or 'anti-coreference' effect. An interesting observation that has often gone unnoticed in the previous literature is that the same generalization holds for languages displaying obligatory indexical shift, suggesting that the two phenomena should be given a uniform account. I argue that this type of 'reporting disjointness' stems from the computation of an antipresupposition at level of person features (Percus 2006, Sauerland 2008a i.a.). Altogether with the aforementioned generalization, the present proposal is shown to be able to correctly derive distribution patterns of both classes of pronouns across languages.

1 Introduction

Some languages can 'shift' indexicals, and use them anaphorically to refer to arguments of the matrix clause. This is exemplified in (1) for Erythrea Tigrinya, a semitic language that allows shifting of first and second person pronouns under verbs of speech:

(1) a. Kidane kə-xɛyəd dɛliɛ ʔallɛxu ʔiłu (nɛyru)
Kidane COMP-IMPF.leave PRF.want.1SG AUX.1SG say.3SG.M AUX.3SG.M
'Kidane$_i$ said that he$_i$ wanted to leave'

b. Kidane kə-xɛyəd dɛliu ʔallo ʔiłu (nɛyru)
Kidane COMP-IMPF.leave PRF.want.3SG.M AUX.3SG.M say.3SG.M AUX.3SG.M
'Kidane$_i$ said that he$_{*i/j}$ wanted to leave' [Tigrinya, personal fieldwork]

This phenomenon, known as indexical shift (henceforth, IS), has been reported for a wide variety of languages pertaining to different families, ranging from Semitic (Amharic, Tigrinya) to Athabaskan (Slave) and Turkic (Uyghur, Chuvash). Languages differ as to which elements undergo shifting: some allow for 1st person shifting only (Slave, Rice 1986), others allow 1st and 2nd person to shift (Uyghur, Shklovsky and Sudo 2014), and some allow for all indexicals to shift without restrictions (Matses, Ludwig et al. 2010; Munro et al. 2012). Languages also vary as to whether IS is obligatory, as in Uyghur (Shklovsky and Sudo, 2014) or Navajo (Speas, 1999), or optional, as in Zazaki (Anand and Nevins 2004; Anand 2006). An interesting observation that has often been overlooked about those languages (a notable exception being Anvari (2020) about Farsi) is that, whenever IS is obligatory, the use of a 3rd person form in speech reports triggers a disjointness inference. The generalization can be stated the following way:

(2) **Disjointness inference in shifty contexts**
In obligatory IS-languages, embedded 3rd person proforms under verbs of speech cannot co-refer with the author of the report.

More precisely, whenever a language L allows for its indexicals to shift under a shifting-licensing predicate, using a 3rd person form in lieu of an indexical will give rise the inference that their

referents are distinct individuals. An early observation of this can be found in Speas (1999) for Navajo, as illustrated in (3):

(3) a. ndoolnish ní
3SG.S.work 3SG.S.say
'He$_i$ says he$_{*i/j}$ will work'

b. ndeeshnish ní
1SG.S.work 3SG.S.say
'He$_i$ says he$_i$ will work'
'He says I will work'

c. nizhdoolnish jiní
4SG.S.work 4SG.S.say
'He$_i$ says he$_i$ will work'

[Navajo, Speas 1999: (3)]

The above data is reminiscent of similar patterns of disjoint reference in languages with logophoric pronouns; in Ewe, for instance, the use of the 3rd person pronoun *e* instead of the logophoric form *yè* indicates that its referent is not the reported speaker, Kofi, but some other, salient male individual:

(4) a. Kofi be yè dzo
Kofi say LOG leave
'Kofi$_i$ said that he$_{i/*j}$ left'

b. Kofi be e dzo
Kofi say 3SG leave
'Kofi$_i$ said that he$_{*i/j}$ left'

[Ewe, Clements 1975]

In (1), the first person indexical can only refer to the matrix subject, and not to the utterance speaker. In order to capture this, a fruitful line of research, pioneered by Anand and Nevins (2004), suggested that the shifting of indexicals may be induced by the presence of a monster operator 👻 in the embedded clause. The semantics of 👻 are straightforward: it rewrites the Kaplanian context coordinates of a context-sensitive expression α-a tuple of parameters consisting of a speaker s, an addressee ad, a world w, a time t and a location l-with the values of the *index*, or circumstances of evaluation, consisting of a similar set of coordinates. Following Anand (2006), I am assuming that indexes are tuples as complex as contexts, not just world-time pairs:

(5) $[\![\text{👻}\ \alpha]\!]^{g,c,i} = [\![\alpha]\!]^{g,i,i}$

Depending on the language, the operator is generally taken to be introduced by attitude verbs such as *say*. Once inserted, all indexicals within its scope will thus inherit the value of the embedded context. An important thing to note is that in most IS languages (including Tigrinya), SIs are unambiguously read *de se* (Schlenker 2003, Anand 2006, Deal 2020 a.o.); sentences involving SIs are judged true only in scenarios in which the reported speaker self-identifies with the attitude holder (Perry 1979, Lewis 1979; see §2.2).

While it allows to capture a wide range of IS-related data crosslinguistically (such as the *shift together* constraint, cf. Anand 2006), the monster-based approach does not predict anything about the disjointness inference alluded to above, that the remaining of this piece will try to cash out in terms of presupposition projection and maximization.

2 Disjointness effects as antipresuppositions

As first observed by Heim (1991) and in a parallel fashion by Hawkins (1991), some utterances involving presupposition triggers seem to be infelicitous in contexts where the truth of a presuppositionally stronger element is entailed, i.e. where the presuppositionally stronger element is common ground:

(6) a. #A moon is bright.
b. The moon is bright.

(7) a. #I broke all my legs running.
b. I broke both legs running.

(6a) presupposes that there is exactly one moon, which is satified in the utterance context; consequently, (6a) will be perceived as odd if uttered. The same goes for (7b), which presuppose that the speaker has exactly two legs; uttering its presuppositionally weaker counterpart

(7a) will trigger the inference that the stronger alternatives do not hold, much like scalar implicatures. Heim (1991) convincingly argues that this kind of inferences cannot readily be analyzed as standard implicatures, because both pairs are equally informative in the given context; she proposes the pragmatic principle *Maximize Presupposition!* to account for the fact that cooperative speakers tend to prefer more informative presuppositional alternatives over their less-informative counterparts. This principle is stated in (8):

(8) **Maximize Presupposition!** (Heim, 1991)
Do not use ϕ if there is a $\psi \in \textsc{Alt}(\phi)$ s.t.
a. $\mathrm{p}(\psi) \subset \mathrm{p}(\phi)$, and
b. $[\![\phi]\!] \equiv [\![\psi]\!]$

Taken as a pragmatic filtering condition on utterances in a given context, *MP!* states that an utterance of ϕ should be avoided if ϕ has an alternative ψ whose presupposition is stronger than that of ϕ, and whose assertive strength (or informativity) are the same in the utterance context. If a competent and cooperative speaker were to utter ϕ under those conditions, then the hearer would consistently infer that she did not utter the presuppositionally stronger ψ on purpose, and that she very likely believes ψ to be false: in other words, the utterance of ϕ would give raise to an *antipresupposition* (Percus, 2006). If *MP!* is a general principle guiding speakers and hearers alike in the interpretation of presuppositions, and if, as we argue below, person features are presupposition triggers, we should expect to observe *MP!*-related effects in the pronominal domain as well.

2.1 The morphosemantics of person

Let us assume that the following features (9) are active across languages, where 1, 2, 3 stand for the persons (Sauerland 2003; McGinnis 2005; Sauerland 2008b; Harbour 2016; Sauerland and Bobaljik 2022). In line with most current research in the semantics of person (Cooper 1983; Heim 2008; Sauerland 2008b; Stokke 2010; Sudo 2012; Charnavel 2019, Sauerland and Bobaljik 2022 a.o.), I consider person features to be interpreted as presuppositions, i.e. partial functions of type $\langle e, e\rangle$ that restrict the domain of interpretation of the expression they are associated with (the pronoun itself being treated as a variable, cf. Heim and Kratzer 1998); since 3rd person pronouns are devoid of person features, no entry is associated with them (Benveniste, 1966), (10).

(9) a. 1: [Part(icipant), Author]
b. 2: [Part(icipant)]
c. 3: [ ]

(10) a. $[\![\ \text{1st}\ ]\!]^{g,c,i} = \lambda x : s(c) \sqsubseteq x.x$
b. $[\![\ \text{2nd}\ ]\!]^{g,c,i} = \lambda x : s(c) \sqsubseteq x \vee a(c) \sqsubseteq x.x$

A key component of this hierarchy is that it is both *asymmetric* (features entail lower ones in the hierarchy) and *nonnovel* (every feature projects onto the higher levels). As (10) illustrate, the set of pronouns of a given language complies with Katzir 2007's definition of structural alternatives: for any given pronoun π, $\textsc{Alt}(\pi)$ is the set of other pronominal forms available in the lexicon of the language, naturally deriving an appropriate scale on which *MP!* can apply, and where features are ranked in terms of semantic markedness (Sauerland 2008b, 2008a): a given feature in the hierarchy will semantically entail all the features below it. The present system predicts that their use are subject to *MP!*; specifically, use of a feature F in the scale will trigger the antipresupposition that stronger, higher-ranked alternatives F' do not hold.

2.2 Antipresuppositions in complex sentences

Indexical shift being mostly an embedded phenomenon, we will need to give an account of person presupposition projection in complex sentences. Following Heim (1992), sentences of the form *x believes that p* or *x says that p* are analyzed as context updates relativized to doxastic alternatives (Hintikka, 1969):

(11) a. $[\![\text{say}]\!]^{g,c,i} = \lambda p.\lambda x.\forall w' \in DOX_{x,w}, p(w') = 1$

b. For any common ground C, C + *John*^{i} *believes he*$_i$ *is hungry* = $\{w \in C : \forall w' \in DOX(j,w), \text{g(i) is hungry in } w'\}$.

This analysis, however, is incomplete, since it fails to deliver two crucial predictions needed in order to capture the data at stake. First, it does not predict any *de se* readings; second, while it captures the meaning of sentences involving doxastic verbs such as *believe* or *think*, it cannot straightforwardly be applied to other predicates such as *say* or *claim*, which quantify over different modal bases (Brasoveanu and Farkas 2007; Anand and Hacquard 2008, 2014). Following Pearson (2015a), I adopt the following entry for *say*-predicates:

(12) $[\![\text{say}]\!]^{g,c,i} = \lambda p.\lambda x.\forall w' \in SAY_{x,w}, p(w') = 1$, where SAY denotes the accessibility relation `compatible-with-what-x-said-in-w`.

With this modification in place, we can now turn to the second issue, namely, the *de se*: while our semantics will correctly delivers *de re* readings of the embedded clause, it will fail to do so in cases where the sentence is true if and only if the agent of *say* recognizes himself as the subject of the attitude introduced by the verb - i.e., has a *de se* attitude about himself. We will modify the entry in (12) in order to allow quantification on properties (and not propositions) of type $\langle e, \langle s, t \rangle \rangle$, where properties are tuples containing both individual and world parameters - *centered propositions*. A centered proposition will be true of an individual relative to a world if and only if this individual correctly self-locates himself in that world; in other words, it will be true if and only if the relevant individual has a *de se* attitude about himself, (13):

(13) **Semantics of *say* (centered worlds version):**
$[\![\text{say}]\!]^{g,c,i} = \lambda P.\lambda x.\forall \langle y, w' \rangle \in SAY_{x,w}, p(w')(y) = 1$

This will ensure that our semantics for attitude verbs, coupled with Heim's mechanism of presupposition projection, will correctly derive *de se* readings of sentences involving both SIs and logophoric pronouns (§3).

With that in place, we can now turn back to disjointness inferences. Consider (1). Recall that, since Tigrinya is an obligatory shifting language, the insertion of 👻 is mandatory under *ʔiłu* '*say*', with the effect that the embedded first person coordinate is shifted towards that of the reported speaker, Kidane. This yields the following pseudo-LF for a sentence like (1b):

(14) Kidane^5 said 👻 I$_5$ want to leave.

Following our semantics as well as Heim's presupposition projection rule, we get the following LF for (14):

(15) $[\![$ Kidane^5 said 👻 I$_5$ want to leave $]\!]^{g,c,i} = 1$ iff $\forall i \in SAY_{w(c)}(Kidane)[s(i) \sqsubseteq (g(5))].\forall i \in SAY_{w(c)}(Kidane)[g(5)$ wants to leave in $w(i)]$.

In words, (14) is true iff the presupposition of the embedded 1st person is satisfied in Kidane's *say*-alternatives, those worlds in which Kidane is the *de se* center in the (shifted) context world $w(i)$, now the world of the index. Consider now an utterance of the following sentence, which is infelicitous if *kidane* and *he* are co-indexed:

(16) Kidane^5 said 👻 he$_5$ wants to leave.

Contrary to its 1st person counterpart, (16) does not presuppose anything. However, in virtue of *Maximize presupposition!* as defined above, its utterance triggers an antipresupposition,

negating all the presuppositional alternatives of the sentence. Consequently, uttering (16) antipresupposes that the speaker does not know whether *g(5)* refers to the speaker or the addressee. However, as noted forcefully in the literature about both scalar implicatures (Spector 2003, Van Rooij and Schulz 2004, Sauerland 2004) and antipresuppositions (Chemla, 2008), this inference is too weak: what the utterance of (16) conveys is actually something stronger, by which the speaker indicates that she does not believe p's presuppositional alternatives to be true. The antipresupposition arrived at with this extra *epistemic step* is indicated in (17), which derives the observed disjointness inference in Tigrinya: *he* and *Kidane* do not co-refer.

(17) **Antipresupposition of 👻 / 3 sentences (with epistemic step):**
a. (16) $\rightsquigarrow CG\neg[\forall i \in SAY_{Kidane,w(c)}[s(i) \wedge a(i) \sqsubseteq (g(5))]$.
b. (16) $\rightsquigarrow$ *g(5)* is neither the reported speaker or addressee.

By way of comparison, consider an utterance of the same sentence in English; it will trigger the same antipresupposition than its Tigrinyan counterpart, with one crucial difference: since no monster is present in the structure, the worlds relative to which the *say*-alternatives of the reported speaker are those accessible from the *actual* context, not those of the index or reported context. As a consequence, the antipresupposition targets the matrix context, deriving the correct disjointness inference for English - *he* and the actual speaker cannot co-refer.

3 Deriving anti-logophoricity

Logophoric languages have a distinct set of 3rd person proforms that are used in embedded contexts under verbs of speech (Clements 1975, Hyman and Comrie 1981, Sells 1987, Culy 1994). Crucially, when used, those pronouns cannot pick up a referent distinct from the reported speaker, or 'logophoric center', in the sense of Sells (1987); using a regular, third person pronoun in the same environment triggers a similar disjointness inference:

(18) a. Oumar Anta inyemɛn waa be gi
Oumar Anta LOG.ACC seen AUX said
'Oumar$_i$ said that Anta had seen him$_i$'
b. Oumar Anta won waa be gi
Oumar Anta 3SG.ACC seen AUX said
'Oumar$_i$ said that Anta had seen him$_{*i/k}$'
[Donno Sɔ, Culy 1994: (1)]

(19) a. #oò kɔ oò dɔ
2SG said 2SG fell
'You$_i$ said you$_i$ fell'
b. oò kɔ oò dɔ-ɛ
2SG said 2SG fell-LOG
'You$_i$ said you$_i$ fell'
[Gokana, Hyman and Comrie 1981: (10)]

Given the similarity between examples (4a), (18), (19) from logophoric languages and example (1) above, it seems reasonable to assume that both phenomena are closely related. In spite of this, logophoric pronouns (LPs) and SIs have been given very different treatments in the literature, for reasons that we cannot review here (Anand 2006; Deal 2020). Before attempting to apply the present analysis to LPs, it might be useful to motivate a unified analysis in the first place. An important similarity concerns the typological distribution of the two classes, which occur in very similar environments - attitude reports. As first proposed by Culy (1994) for LPs and by Deal (2017) for SIs, both classes across languages seem to be licensed by a common range of predicates, organized into an implicational scale; second, both enforce *de se* interpretations (Schlenker 2003; Adesola 2006; Anand 2006; Haida 2009; Sudo 2012; Deal 2020; Bimpeh et al. 2022). This suggests that a unified account of both phenomena is worth pursuing.

The behavior of LPs can straightforwardly be captured with our antipresupposition account, with further assumptions about logophoric systems. In order to consistently maintain the required featural asymmetry at the level of alternatives, let us assume that, instead of being

specified with a dedicated feature [Log], as in most accounts (Schlenker 1999, von Stechow 2002, 2003, and more recently Bimpeh et al. 2022), logophors lack a feature that first person pronominals have: a feature [Actual], which 1st person pronouns come specified with (the [Actual] feature is inspired from Schlenker (2003), and closely resembles his [+ Author*]). I assume that logophoric languages like Ewe make use of the following feature set:

(20) **Features of logophoric systems**
   a. 1: [Author], [Participant], [Actual]
   b. Log: [Author], [Participant]
   c. 2: [Participant]
   d. 3: [ ]

(21) **Semantics of logophoric features:**
   a. $[\![1]\!]^{g,c,i} = \lambda x : s(c) \sqsubseteq x.x$
   b. $[\![\text{Log}]\!]^{g,c,i} = \lambda x : s(c) \sqsubseteq x \vee s(i) \sqsubseteq x.x$
   c. $[\![2]\!]^{g,c,i} = \lambda x : s(c) \sqsubseteq x \vee a(c) \sqsubseteq x \vee s(i) \sqsubseteq x \vee a(i) \sqsubseteq x.x$

Once such a hierarchy is assumed, deriving disjointness inferences triggered by ordinary 3rd person pro-forms in LP-languages becomes straightforward. Take example (4b) above. What we observe is that the use of a standard 3rd person form in the logophoric environment introduced by *be 'say'* suggests that the two referents are distinct individuals. This is because the choice of (4b) over its logically stronger counterpart (4a) triggers our familiar inference that *Kofi* refers neither to the author nor addressee of either the utterance or the reported context. As a consequence, it is assumed that *Kofi* and e must denote distinct individuals.

(22) Kofi[5] said he$_5$ wanted to leave.

(23) **Antipresupposition of 3 / Log sentences (with epistemic step):**
   a. (22) $\rightsquigarrow CG \neg[\forall i \in SAY_{Kofi,w(c)}[s(c) \wedge a(c) \wedge s(i) \wedge a(i) \sqsubseteq (g(5))]]$.
   b. $\rightsquigarrow$ *g(5)* is neither the actual nor reported speaker or addressee.

The present theory allows to capture further data. As noted by Hyman and Comrie (1981) for Gokana, logophoric pronouns cannot take 1st person pronouns as antecedents. In other words, for a given speech report, when the reported and current speaker are one and the same individual, a logophor cannot be used:

(24) a. mm kɔ mm dɔ
   1sg said 1sg fell
   'I$_i$ said I$_i$ fell'

   b. #mm kɔ mm dɔ-ɛ
   1sg said 1sg fell-log
   'I$_i$ said I$_i$ fell'
   [Gokana, Hyman and Comrie 1981: (11)]

Similar data are reported for Wan (Niger-Congo, Ivory Coast; Nikitina 2012), Ewe (Pearson, 2015b) and Danyi Ewe (Niger-Congo, Togo; O'Neill 2015), as well as Ibibio (Niger-Congo, Southern Nigeria; Newkirk 2017). This pattern is correctly predicted by the antipresupposition account, given the asymmetrical hierarchy of features posited in (20): in cases where the antecedent is first person and refers to the current speaker, a first person must be used in the embedded sentence. Since the feature set of 1st person is semantically more marked than that of Log, then it ensues that $[\![\text{1st}]\!] \subset [\![\text{Log}]\!]$; any utterance of Log will trigger the antipresupposition in (23), which contradicts the matrix use of the 1st person, where the actual speaker intends to refer to himself with that form. Another interesting typological fact that our theory can derive concerns the use of 2nd person in logophoric environments. In some languages, logophoric contexts exhibit a special case of 'person neutralization' between third and second person; as a consequence, logophoric pronouns can take second person antecedents as well as third, with singular and plural number features alike (cf. example (19)). As already discussed, first person antecedence is excluded. On the present account, varieties of person neutralization exemplified in (19) are correctly predicted: it is expected that a sentence where the author of

the embedded speech event is referred to using a 2nd person pronoun will be infelicitous, regardless of what his discourse status in the actual context is; a logophor should be used instead - which is just what we observe. In these, the 2nd person and the logophor are coreferential, the addressee of the utterance context also being the reported speaker. When these coincide, a 2nd person cannot be used on pains of triggering a disjointness inference, as in (19a). Last, the present system makes one further prediction: due to their relative underspecification compared to LPs, 2nd person pronouns should be able to refer to reported addressees, a prediction that seems borne out, cf. (25):

(25) è gé zò ɓé **là** **ɓà** pólì
3SG said come then 2SG LOG.SG wash
'She$_i$ said come and wash me$_i$.' [Wan, Nikitina 2012: (18)]

The data above is quite interesting when compared to IS-systems discussed in the previous section, since the second person in (25) is 'shifty' in a similar sense. What this suggests is that logophoric languages have somehow grammaticalized a version of the monster operator 👻 that shifts the embedded speaker coordinates, allowing reference to embedded speech context authors to be lexicalized by a dedicated logophoric form, thereby confining the 1st person to genuine indexical uses (an analysis along these lines is in fact suggested for Wan by Nikitina 2020).

References

Adesola, O. P. (2006). A-bar dependencies in the yoruba reference-tracking system. *Lingua*, 116(12):2068–2106.

Anand, P. (2006). *De De Se*. PhD thesis, MIT.

Anand, P. and Hacquard, V. (2008). Epistemics with attitude. In *Semantics and linguistic theory*, volume 18, pages 37–54.

Anand, P. and Hacquard, V. (2014). Factivity, belief and discourse. *The art and craft of semantics: A festschrift for Irene Heim*, 1:69–90.

Anand, P. and Nevins, A. (2004). Shifty operators in changing contexts. In *Semantics and Linguistic Theory*, volume 14, pages 20–37.

Anvari, A. (2020). *Meaning in context*. PhD thesis, Ecole Normale Supérieure.

Benveniste, E. (1966). La nature des pronoms. In *Problèmes de linguistique générale*. Gallimard.

Bimpeh, A. A., Driemel, I., Bassi, I., and Silleresi, S. (2022). Obligatory de se logophors in ewe, yoruba and igbo: variation and competition. In *Talk given at WCCFL 40, May 13th, 2022*.

Brasoveanu, A. and Farkas, D. (2007). Say reports, assertion events and meaning dimensions. *Pitar Mos: A Building with a View. Papers in Honour of Alexandra Cornilescu, Editura Universitatii din Bucuresti, Bucharest*, pages 175–196.

Charnavel, I. (2019). Presupposition failure and intended pronominal reference: Person is not so different from gender after all. *Linguistic Inquiry*, 50(2):372–387.

Chemla, E. (2008). An epistemic step for anti-presuppositions. *Journal of Semantics*, 25(2):141–173.

Clements, G. N. (1975). The logophoric pronoun in ewe: Its role in discourse. *Journal of West African Languages*, 10:141–177.

Cooper, R. (1983). *Quantification and syntactic theory*, volume 21. Springer Science & Business Media.

Culy, C. (1994). Aspects of logophoric marking. *Linguistics*.

Deal, A. R. (2017). Shifty asymmetries: universals and variation in shifty indexicality. *Ms., University of California, Berkeley*.

Deal, A. R. (2020). *A Theory of Indexical Shift*. The MIT Press.

Haida, A. (2009). Logophoricity in tangale. In *NELS 40 Semantics Workshop on Pronouns*.

Harbour, D. (2016). *Impossible persons*. Mit Press.

Hawkins, J. A. (1991). On (in) definite articles: implicatures and (un) grammaticality prediction. *Journal of linguistics*, 27(2):405–442.

Heim, I. (1991). Artikel und definitheit. In *Semantik: Ein Internationales Handbuch*, page 487–535. Berlin: de Gruyter.

Heim, I. (1992). Presupposition projection and the semantics of attitude verbs. *Journal of semantics*, 9(3):183–221.

Heim, I. (2008). Features on bound pronouns. In Harbour, D., Adger, D., and Béjar, S., editors, *Phi theory: Phi-features across modules and interfaces*. Oxford University Press.

Heim, I. and Kratzer, A. (1998). *Semantics in generative grammar*, volume 13. Blackwell Oxford.

Hintikka, J. (1969). Semantics for propositional attitudes. In *Models for modalities*, pages 87–111. Springer.

Hyman, L. M. and Comrie, B. (1981). Logophoric reference in gokana. *Journal of African Languages and Linguistics 3: 19-37*.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy*, 30(6):669–690.

Lewis, D. (1979). Attitudes de dicto and de se. *The philosophical review*, 88(4):513–543.

Ludwig, R. A., Munro, R., Fleck, D. W., and Sauerland, U. (2010). Reported speech in matses: Obligatory perspective shift with syntactic transparency. *Proceedings of SULA*, 5:33–47.

McGinnis, M. (2005). On markedness asymmetries in person and number. *Language*, 81(3):699–718.

Munro, R., Ludwig, R., Sauerland, U., and Fleck, D. W. (2012). Reported speech in matses: Perspective persistence and evidential narratives. *International Journal of American Linguistics*, 78(1):41–75.

Newkirk, L. (2017). Logophoricity in ibibio. *Handout of talk given at ACAL 48, Rutgers University, March 2017. https://lydianewkirk.com/research/ACAL3-17Handout.pdf*. https://lydianewkirk.com/research/ACAL3-17Handout.pdf.

Nikitina, T. (2012). Logophoric discourse and first person reporting in wan (west africa). *Anthropological Linguistics*, 54(3):280–301.

Nikitina, T. (2020). Logophoricity and shifts of perspective: New facts and a new account. *Functions of Language*, 27(1):78–99.

O'Neill, T. (2015). The distribution of the danyi ewe logophor yi. *Talk given at the LSA Annual Meeting*.

Pearson, H. (2015a). Attitude verbs. *Ms., Queen Mary University of London*.

Pearson, H. (2015b). The interpretation of the logophoric pronoun in ewe. *Natural Language Semantics*, 23(2):77–118.

Percus, O. (2006). Antipresuppositions. *Theoretical and empirical studies of reference and anaphora*, 52:73.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, pages 3–21.

Rice, K. (1986). Some remarks on direct and indirect discourse in Slave (Northern Athapaskan). In Coulmas, F., editor, *Direct and indirect speech*, volume 31, page 47. De Gruyter Mouton.

Sauerland, U. (2003). A new semantics for number. In Young, R. B. and Zhou, Y., editors, *Proceedings of SALT 13*, pages 258–275. CornellUniversity, Ithaca, N.Y.: CLC-Publications.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and philosophy*, 27(3):367–391.

Sauerland, U. (2008a). Implicated presuppositions. In *The discourse potential of underspecified structures*, volume 8, pages 581–600. Walter de Gruyter Berlin.

Sauerland, U. (2008b). On the semantic markedness of phi-features. In Harbour, D., Adger, D., and Béjar, S., editors, *Phi theory: Phi-features across modules and interfaces*. Oxford University Press.

Sauerland, U. and Bobaljik, J. D. (2022). Cumulative conjunction and exhaustification in morphology. *Ms., ZAS & Harvard University*.

Schlenker, P. (1999). *Propositional attitudes and indexicality: a cross categorial approach*. PhD thesis, Massachusetts Institute of Technology.

Schlenker, P. (2003). A plea for monsters. *Linguistics and philosophy*, 26(1):29–120.

Sells, P. (1987). Aspects of logophoricity. *Linguistic Inquiry*, 18(3):445–479.

Shklovsky, K. and Sudo, Y. (2014). The syntax of monsters. *Linguistic Inquiry*, 45(3):381–402.

Speas, M. (1999). Person and point of view in navajo direct discourse complements. University of Massachhussets Amherst.

Spector, B. (2003). Scalar implicatures: Exhaustivity and gricean reasoning. In ten Caten, B., editor, *Proceedings of the Eighth ESSLLI Student Session*, page 277–88. Vienna: FOLLI.

Stokke, A. (2010). *Indexicality and presupposition: explorations beyond truth-conditional information*. PhD thesis, University of St Andrews.

Sudo, Y. (2012). *On the semantics of phi features on pronouns*. PhD thesis, Massachusetts Institute of Technology.

Van Rooij, R. and Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information*, 13(4):491–519.

von Stechow, A. (2002). Binding by verbs: Tense, person and mood under attitudes. In *Proceedings of NELS*, volume 33, pages 379–403.

von Stechow, A. (2003). Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In *Proceedings of NELS*, volume 33, pages 379–404.

The semantics of Rudin constructions in Romanian*

Ivano Caponigro[1] and Anamaria Fălăuș[2]

[1] University of California San Diego, U.S.A
ivano@ucsd.edu
[2] Centre National de la Recherche Scientifique-Laboratoire de Linguistique de Nantes, France
anamaria.falaus@cnrs.fr

Abstract

We broaden the empirical and formal landscape of *wh*-constructions by describing and analyzing an understudied non-interrogative *wh*-construction requiring multiple *wh*-expressions, to which we refer as the 'Rudin construction'. We argue that, unlike other multiple *wh*-constructions studied to date, Rudin constructions are biclausal and should be semantically analyzed as establishing identity between the extensions of two relations.

1 Introducing the Rudin construction

Four Romanian friends decide to have a potluck picnic on Sunday. Unfortunately, the friends haven't shared their food restrictions or preferences ahead. Once they get together, they find out that none of them likes or can eat any of the food the others prepared. So, each of them ends up eating only their own food. The not-so successful end of this potluck story can be precisely and concisely conveyed in Romanian by means of the sentence in (1).

(1) La picnicul de duminică _ a mâncat _ [cine$_1$ ce mâncare$_2$ _$_1$ a pregătit _$_2$].
at picnic-the of Sunday has eaten who what food has prepared
'Each person ate at the picnic on Sunday what (s)he prepared, and nothing else.'

The sentence in (1) looks like it is made of two fully tensed clauses, each with a fully inflected transitive verb ('eat', 'prepare'). Both predicates are obligatorily missing their subject and object arguments—highlighted with underscores. If either argument is realized in either clause, the whole sentence becomes fully unacceptable. The bracketed clause is introduced by two (underlined) *wh*-expressions (i.e., *wh*-words or *wh*-phrases) that are linked to its missing subject and object—highlighted with the shared subscripts "1" and "2". The other clause doesn't have any overt clause-internal marker correlating with its missing subject or object—highlighted with plain underscores without subscripts. The whole sentence is interpreted as asserting that each eater ate at the picnic on Sunday all and only the food that (s)he prepared. In other words, the sentence asserts the identity between the set of ordered pairs of ⟨eater, eaten-food⟩ associated with the first clause and the set of ordered pairs ⟨food-preparer, prepared-food⟩ associated with the second clause.

The sentence in (1) is an example of a typologically rare construction that, on the other hand, is extremely productive in Romanian. Other examples are given in (2) with a missing object (a locative object in (2b)) and a missing adjunct in each example, as well as in (3) with three missing constituents and corresponding *wh*-expressions.

(2) a. Fac [ce cum îmi place].
do.1SG what how me.DAT likes
'I do whatever I like however I like it.'

b. Mergi [unde când vrei]!
go.IMP.2SG where when want.2SG
'Go wherever you want whenever you want!'

*We would like to thank the reviewers for their helpful comments and suggestions. We are solely responsible for any remaining errors. The names of the authors are listed in alphabetical order.

(3) Gustă [cine ce când termină de preparat].
tastes who what when finishes of prepared
'Everyone tastes what they are done preparing when they are done preparing it.'

We call the construction in (1)–(3) a "Rudin construction" since Catherine Rudin was the first scholar to mention and describe it, to the best of our knowledge (Rudin 1986, 2007, 2008). It has been largely neglected since with the exception of the brief discussion in Caponigro and Fălăuş (2020) and the semantic analysis in Nicolae (2020). Rudin constructions seem to be attested in other Balkan languages like Bulgarian or Macedonian, but at this stage we focus on Romanian, leaving it open whether the constructions in those languages are identical. In this paper, we propose a compositional semantic analysis that captures the semantic properties of Rudin constructions like (1)–(3) and argue that it is superior to the one in Nicolae (2020).

2 Semantic analysis of the Rudin construction

A Rudin construction is always a fully biclausal construction, as schematized in (4).

(4) $[_{\text{Clause1}} \cdots __1 \cdots __2 \cdots __n \cdots \; [_{\text{Clause2}} \; \textbf{wh-}_1 \; \textbf{wh-}_2 \; \textbf{wh-}_n \cdots __1 \cdots __2 \cdots __n \cdots]]$, $n \geq 0$

Clause_1 of a Rudin construction always occurs first (left-most) and behaves like the main clause, determining the semantic and pragmatic features of the whole Rudin construction. If Clause_1 is declarative, then the whole Rudin construction is declarative, as in (1); if Clause_1 is interrogative, as in (5), or imperative (2b), then the whole Rudin construction will be interrogative or imperative, respectively. Clause_1 allows for left-dislocated topicalized constituents or moved *wh*-constituents—the latter case is illustrated in (5).

(5) A: **Unde** a mâncat (duminică) [cine ce mâncare a pregătit]? B: La picnic.
where has eaten Sunday who what food has prepared at picnic
'Where did (on Sunday) everyone eat the food they prepared?' 'At the picnic.'

Clause_2 of a Rudin construction always occurs to the right edge of Clause_1. We tentatively assume that Clause_2 is an embedded *wh*-clause (a CP) that is adjoined somewhere in the left periphery of Clause_1, the main clause (we leave further syntactic investigation to future work). Clause_1 and Clause_2 are missing the same number and kinds of constituents (arguments or adjuncts)—at least two. While Clause_1 leaves the missing constituents unmarked, Clause_2 is obligatorily introduced by a series of fronted *wh*-expressions matching the missing arguments or adjuncts. Clause_2 can make use of all the *wh*-expressions attested in interrogative clauses and exhibits similar ordering constraints. The whole Rudin construction can be embedded within a larger sentence, as shown in (6).

(6) Mă întreb dacă [la picnicul de duminică a mâncat cine ce a pregătit].
me wonder.1SG if at picnic-the of Sunday has eaten who what has prepared
'I wonder whether at the picnic on Sunday everyone ate what they prepared.'

In this paper, we focus on developing a semantic analysis for Rudin constructions. In particular, we argue that a Rudin construction denotes identity between the extensions at the world of evaluation w_0 of two $2{+}n$-place relations ($n \geq 0$) ranging over ordered $2{+}n$-tuples—one relation being denoted by the relevant portion of Clause_1[1], the other being denoted by the whole Clause_2,

[1]The "relevant portion" of Clause_1 is the part of the clause going all the way up to the node right below the node to which Clause_2 adjoins. Henceforth *Clause*$_1$ is used as a shortcut for that portion of Clause_1.

as schematized in (7). The "2+n" requirement captures the fact that a Rudin construction always has at least two missing constituents.

(7) $[\![\lambda x_1 \lambda x_2 \ldots \lambda x_n \textbf{Clause}_1(x_1, x_2 \ldots x_n)]\!]^{w0} = [\![\lambda x_1 \lambda x_2 \ldots \lambda x_n \textbf{Clause}_2(x_1, x_2 \ldots x_n)]\!]^{w0}$

We propose that the two 2+n-place relations in (7) are derived by means of the same basic syntax/semantics mechanisms: the operator-variable relation that has been assumed for several constructions that display overt or covert operator movement. Clause_2 is more "transparent" in this respect: the overt, fronted *wh*-expressions in Spec,CP signal the missing constituents within the lower TP and license *wh*-traces/variables in their base-generated position. Each moved *wh*-constituent triggers (type-flexible) λ-abstraction over the variable it licenses right before combining with its sister (i.e., the portion of Clause_2 the given *wh*-expression c-commands). Therefore, if the portion of Clause_2 that is c-commanded by a given *wh*-expression denotes a semantic object of type m and the *wh*-variable is of type r, the resulting constituent after λ-abstraction (i.e., the sister of the *wh*-expression) denotes a semantic object of type $\langle rm \rangle$. We assume flexible semantic types for *wh*-expressions as well: if its sister is of type $\langle rm \rangle$, then the *wh*-expression will be of type $\langle rm, rm \rangle$. The *wh*-expression semantically acts as a restrictor: it applies to an object of type $\langle rm \rangle$ to return an object of the same type with further semantic restrictions like 'human' if the *wh*-expression is 'who', inanimate if it is 'what', etc. For instance, Clause_2 in (1) ends up denoting a 2-place relation in (8e) by first saturating the 2-place predicate 'prepare' with the *wh*-traces/variables in object and subject positions (8a) and then abstracting over each of those *wh*-variables, one at a time (8b,d), right before each of them combines with the corresponding *wh*-expression (8c,e).

(8) a. $[\ __1$ a pregătit $__2\ ] \rightsquigarrow \text{prepared}(x_1, x_2)_t$
b. $[\lambda_2[\ __1$ a pregătit $__2\ ]\ ] \rightsquigarrow \lambda x_2 \text{prepared}(x_1, x_2)_{\langle et \rangle}$
c. $[\underline{\text{ce mâncare}_2}\ \lambda_2[\ __1$ a pregătit $__2\ ]\ ] \rightsquigarrow \lambda x_2[\text{prepared}(x_1, x_2) \wedge \text{food}(x_2)]_{\langle et \rangle}$
d. $[\lambda_1[\underline{\text{ce mâncare}_2}\ \lambda_2[\ __1$ a pregătit $__2\ ]]] \rightsquigarrow \lambda x_1 \lambda x_2[\text{prepared}(x_1, x_2) \wedge \text{food}(x_2)]_{\langle e,et \rangle}$
e. $[\underline{\text{cine}_1}\ \lambda_1[\underline{\text{ce mâncare}_2}\ \lambda_2[\ __1$ a pregătit $__2\ ]]] \rightsquigarrow \lambda x_1 \lambda x_2[\text{prepared}(x_1, x_2) \wedge \text{food}(x_2) \wedge \text{human}(x_1)]_{\langle e,et \rangle}$

We envision that the same derivational mechanism is at play to derive the denotation of Clause_1 as a 2+n-place relation. The only difference is that we assume phonologically null operators (Op) as silent counterparts of *wh*-expressions, since the missing constituents of Clause_1 are not marked by any *wh*-expressions or any other marker. Silent operators too move to the left periphery, license traces/variables in their base-generated positions and then bind those traces/variables via λ-abstraction. Notice that something along these lines is usually assumed to handle relative clauses with no overt relative pronoun or relative markers, like *the food they prepared*, *the way he laughs*, or *the day she left* in English. For instance, the semantic derivation of Clause_1 in (1) would proceed as in (9a–e), which parallels the derivation of Clause_2 in (8a–e). The only main difference is that silent operators do not carry any further semantic restrictions, unlike overt *wh*-expressions, and, therefore, semantically behave like identity functions. In other words, they are semantically inert and their main function is to license *wh*-variables and trigger λ-abstraction (this is why the logical translations in (9b,c) and (9d,e) are identical).

(9) a. $[\ __1$ a mâncat $__2\ ] \rightsquigarrow \text{ate}(x_1, x_2)_t$
b. $[\lambda_2[\ __1$ a mâncat $__2\ ]] \rightsquigarrow \lambda x_2 \text{ate}(x_1, x_2)_{\langle et \rangle}$
c. $[\textbf{Op}_2\ \lambda_2[\ __1$ a mâncat $__2\ ]] \rightsquigarrow \lambda x_2 \text{ate}(x_1, x_2)_{\langle et \rangle}$
d. $[\lambda_1[\textbf{Op}_2\ \lambda_2[\ __1$ a mâncat $__2\ ]]] \rightsquigarrow \lambda x_1 \lambda x_2 \text{ate}(x_1, x_2)_{\langle e,et \rangle}$
e. $[\textbf{Op}_1\ \lambda_1[\textbf{Op}_2\ \lambda_2[\ __1$ a mâncat $__2\ ]]] \rightsquigarrow \lambda x_1 \lambda x_2 \text{ate}(x_1, x_2)_{\langle e,et \rangle}$

When Clause_2 adjoins to Clause_1, we propose that their denotations at the world of evaluation are identified by means of the new operator in (10), which we label 'Rudin operator' (O_{RUDIN}). The Rudin operator takes the $2{+}n$ relation R of type m denoted by Clause_2 as its first argument and the $2{+}n$ relation Q of the same type m denoted by Clause_1 as its second argument to return the truth if those relations are extensionally identical at the world of evaluation.

(10) *Rudin Operator*: $\lambda \text{R}_{2+n} \lambda \text{Q}_{2+n} [\text{R}_m = \text{Q}_m]_{\langle m, mt \rangle}$

The Rudin operator contains a high degree of flexibility, as the direct reflex of the variable number and nature of missing constituents within a Rudin construction. This operator is flexible as far as its semantic type is concerned, as highlighted by the unspecified type m inside the Rudin operator semantic type $\langle m, mt \rangle$ in (10). This flexibility follows from the variable n-ary of the relations the Rudin operator identifies: n-ary of two or more, as long as identical between the two relations. This captures the fact that a Rudin construction can have two or more missing constituents, as long as the number of missing constituents is the same between Clause_1 and Clause_2. The flexibility of the semantic type of the Rudin operator also depends on the variable nature (i.e., semantic type) of the arguments of its relations, as long as they are the same across the two relations. This reflects the fact that the semantic type of a missing constituent that is marked by the *wh*-expression 'who' may not be the same as the missing constituent that is signaled by the *wh*-expression 'where' or 'when'.[2] Finally, the Rudin operator makes use of a flexible relation of identity, which applies to objects of variable semantic type. Syntactically, the Rudin operator occupies the head of the highest functional projection of Clause_2, as shown in (11).

(11) $[_{\text{CP-Clause1}} \ldots [_{\text{CP-Clause2}} \ \mathbf{O_{RUDIN}} \ [_{\text{CP-Clause2}} \ldots]]]$

Let's exemplify how the Rudin operator works by returning to the semantic derivation of (1). First, we apply the Rudin operator as defined in (10) to Clause_2 in (1), whose logical translation was given in (8e). This is shown in (12). Then, we apply (12) to Clause_1, whose logical translation was given in (9e). The final result is the identity in (13).

(12) $[\mathbf{O_{RUDIN}} \ [\underline{\text{cine}}_1 \ \lambda_1 [\underline{\text{ce mâncare}}_2 \ \lambda_2 [__1 \text{ a pregătit } __2 \ ]]]] \rightsquigarrow$
$\lambda \text{R} \lambda \text{Q} [\text{R}_{\langle e,et \rangle} = \text{Q}_{\langle e,et \rangle}] (\lambda \text{x}_1 \lambda \text{x}_2 [\text{prepared}(\text{x}_1,\text{x}_2) \wedge \text{food}(\text{x}_2) \wedge \text{human}(\text{x}_1)])$
$\lambda \text{Q} [\lambda \text{x}_1 \lambda \text{x}_2 [\text{prepared}(\text{x}_1,\text{x}_2) \wedge \text{food}(\text{x}_2) \wedge \text{human}(\text{x}_1)] = \text{Q}]$ (*by λ-reduction*)

(13) $[[\underline{\text{Op}}_1 \ \lambda_1 [\underline{\text{Op}}_2 \ \lambda_2 [__1 \text{ a mâncat } __2]]] \ [\mathbf{O_{RUDIN}} [\underline{\text{cine}}_1 \ \lambda_1 [\underline{\text{ce mâncare}}_2 \ \lambda_2 [__1 \text{ a pregătit } __2]]]]]$
$\rightsquigarrow \lambda \text{Q} [\lambda \text{x}_1 \lambda \text{x}_2 [\text{prepared}(\text{x}_1,\text{x}_2) \wedge \text{food}(\text{x}_2) \wedge \text{human}(\text{x}_1)] = \text{Q}] (\lambda \text{x}_1 \lambda \text{x}_2 \text{ate}(\text{x}_1,\text{x}_2))$
$\lambda \text{x}_1 \lambda \text{x}_2 [\text{prepared}(\text{x}_1,\text{x}_2) \wedge \text{food}(\text{x}_2) \wedge \text{human}(\text{x}_1)] = \lambda \text{x}_1 \lambda \text{x}_2 \text{ate}(\text{x}_1,\text{x}_2)$ (*by λ-reduction*)

The Rudin operator also provides an answer to the question of what ensures the correct number and nature of the missing constituents in Clause_1. (Remember that the missing constituents in Clause_1 are not marked, unlike those in Clause_2.) Let's assume that the grammar is free to generate clauses with any number and any kind of missing arguments and/or adjuncts without any overt marking, as in Clause_1. The Rudin operator requires its two clausal arguments to be of the same semantic type: if Clause_1 had a different number or type of missing constituents than Clause_2, then the two clauses would not have the same semantic type, the Rudin operator would fail to apply, and the semantic derivation of the whole Rudin construction would crash.

Another welcome consequence of our analysis is that it provides an account for the "impression" that each *wh*-expression in a Rudin construction is "connected" to both Clause_1 and Clause_2—a feature that sets the Rudin construction apart from other kinds of multiple

[2] On the identity of semantic types among *wh*-expressions, see Caponigro and Pearl (2008, 2009).

wh-clauses in Romanian (see Caponigro and Fălăuş 2020). According to our analysis, the *wh*-expressions per se are part of Clause_2 only and license the missing constituents only in Clause_2. Still, as discussed above, *wh*-expressions affect the semantic type of Clause_2, which in turn affects the semantic type of the Rudin operator, which in turn requires Clause_1 to have the same semantic type and, therefore, the same number and kinds of missing constituents as Clause_2.

A further welcome consequence of our account is that it captures the intuition that the interpretation of *wh*-expressions in a Rudin construction is akin to a universal quantifier like 'every one' or 'each one' or a free choice item like 'anyone' or 'whoever', as highlighted in the English translations of our examples of the Rudin construction. This intuition is strengthened by native speakers' willingness to paraphrase a Rudin construction with sentences containing overt universal quantifiers or free choice items as in (14).

(14) Mănâncă {**fiecare**/ **oricine**} **(ori)ce** aduce.
eats everyone/ anyone what(ever) brings
'Everyone eats what(ever) they bring.'

Our analysis accounts for this intuition without assuming any universal quantifier, but by means of the identity between the extensions of two relations that is imposed by the Rudin operator.

Last, our analysis crucially relies on flexible λ-abstraction and identity between extensions of variable n–ary relations. These are the same tools that are used in the "classical" semantic analysis of *wh*-interrogative clauses as partitions in Groenendijk and Stokhof (1982).

3 Comparison with a previous account

The Rudin construction has received little attention in general, let alone within formal semantics. We are aware of only one other semantic analysis—the one in Nicolae (2020).[3] In this section, we briefly summarize it and discuss additional facts in favor of our proposal.

Nicolae (2020) discusses Rudin constructions with two missing arguments (suggesting that missing adjuncts may be treated as arguments as well). Within this restricted empirical landscape, she analyzes both Clause_1 and Clause_2 as denoting sets of functions of type $\langle ee \rangle$, although their semantic compositions are partially different. The denotation of Clause_1 is derived by turning the lexical entry of its transitive predicate from a semantic object of type $\langle e, et \rangle$ to one of type $\langle ee, t \rangle$ by means of the type-shifter TSH. For instance, Clause_1 in (1) would be initially analyzed as in (15a), to which TSH would apply resulting in (15b).

(15) a. $\text{ate} \rightsquigarrow \lambda y_e \lambda x_e \text{ate}(x,y)$ b. $\text{TSH}(\text{ate}) \rightsquigarrow \lambda f_{\langle ee \rangle} \forall x[\text{ate}(x,f(x))]$

The semantic derivation of Clause_2, instead, builds on the analysis of multiple correlative clauses in Dayal (1996), and the treatment of functional multiple *wh*-interrogative clauses in Abels and Dayal (2017) and Xiang (2021). Clause_2 denotes a set of functions of type $\langle ee \rangle$, which for Clause_2 in (1) amounts to the set of functions from people to food they prepared, as in (16).

(16) $[\text{who what food prepared}] \rightsquigarrow \lambda f_{\langle ee \rangle}[\text{Range}(f)=\text{food} \wedge \forall x[\text{human}(x) \rightarrow \text{prepared}(x,f(x))]]$

The semantic contributions of the two clauses are combined by means of the 2-place operator THE, defined as in (17).

[3] Nicolae (2020) follows Rudin (1986) in calling what we label Rudin constructions "multiple *wh* relative clauses". We have shown elsewhere that Romanian allows for true free relative clauses with multiple *wh*-expressions, which exhibit clear differences with Rudin constructions (Caponigro and Fălăuş 2020).

(17) THE $\rightsquigarrow \lambda F_{\langle ee,t\rangle}\lambda G_{\langle ee,t\rangle}\exists f_{\langle ee\rangle}[(f{=}\iota g \text{ s.t. } F(g)) \wedge G(f)]$

THE first applies to the denotation of Clause$_2$ in (16) and the resulting predicate then applies to the denotation of Clause$_1$ in (15b). The whole sentence in (1) receives the logical translation in (18), which is true iff there is a unique function from individuals to the food they prepared and everyone ate the thing they prepared.

(18) [THE(who what food prepared)](ate) $\rightsquigarrow$
$\exists f[(f{=}\iota g \text{ s.t. } \text{Range}(g){=}\text{food} \wedge \forall x[\text{human}(x) \rightarrow \text{prepared}(x,g(x))]) \wedge \forall x[\text{ate}(x,f(x))]]$

A first crucial difference between our analysis and Nicolae's has to do with the truth conditions of Rudin constructions, as also noted by one of our reviewers. Specifically, the account in Nicolae (2020) requires there to be only one function $\langle ee\rangle$ associated with Clause$_2$, while no such restriction applies to the functions associated with Clause$_1$. In the case of (1), this implies that there's only one mapping between people and the food they prepared and there can be more than one mapping between people and the food they ate. One of these mappings would be identical to the one associated with Clause$_1$, that is, it would map people with the food they prepared and ate, but there could be another mapping according to which people also ate food other than the one they prepared. This is contrary to the intuitions reported by our consultants for the Rudin construction in (1), whereby each person eats only the things (s)he prepared. Our semantic analysis, which requires identity between the set of ⟨eater, eaten-food⟩ pairs and the set of ⟨food-preparer, prepared-food⟩ pairs, delivers the correct interpretation.

Another significant difference between our analysis and Nicolae's has to do with the number of missing constituents in Rudin constructions. As already mentioned, Nicolae (2020) is specifically couched for Rudin constructions with only two missing constituents, while there are Rudin constructions with three (or more), as shown in (3). The crucial components of our analysis–flexible λ-abstraction and flexible Rudin operator–have been designed with this empirical fact in mind. Nicolae's analysis, instead, crucially builds on a functional dependency between the higher and the lower *wh*-constituents; it's unclear how a third *wh*-constituent would be handled. Also, both the type-shifter TSH and the 2-place operator THE are inherently defined for two variables rather than three (or more). It remains to be seen how to broaden these formal tools so as to handle Rudin constructions with three (or more) missing constituents.

4 Conclusions and outstanding issues

We have shown that Romanian has a biclausal non-interrogative multiple *wh*-construction—the Rudin construction—and have argued that it should be semantically analyzed as establishing identity between the extension of two relations. The relations involved in Rudin constructions are always $2{+}n$-place relations with $n \geq 0$. In other words, no Rudin construction is allowed with just one missing constituent for each clause (evidence supporting this generalization needs to be omitted here for reasons of space). The outstanding issue that we leave for future investigation is why Rudin constructions exhibit this restriction. No component of our analysis predicts that. The analysis in Nicolae (2020) doesn't provide a principled explanation of this restriction either. In fact, it is expressly based on analyses of correlative and interrogative clauses—constructions that easily allow for only one *wh*-expression/missing constituent.

We have focused on Romanian, but Rudin constructions seem to be attested in other languages as well (Rudin 1986, 2007, 2008). A crosslinguistic investigation may provide a better grasp of generalizations and properties concerning Rudin constructions and further develop their syntactic and semantic analyses. We hope to have started contributing to this enterprise.

References

Abels, Klaus and Veneeta Dayal (2017). "On the syntax of multiple sluicing". In: *Proceedings of NELS 47*. GLSA publications.

Caponigro, Ivano and Anamaria Fălăuş (2020). "Unveiling Multiple *wh*-Free Relative Clauses and their functional *wh*-words". In: *Semantics and Pragmatics* 13.2.

Caponigro, Ivano and Lisa Pearl (2008). "Silent prepositions: Evidence from free relatives". In: *The syntax and semantics of spatial P*. Ed. by Anna Asbury et al. Amsterdam: Benjamins, pp. 365–385.

Caponigro, Ivano and Lisa Pearl (2009). "The nominal nature of *where*, *when*, and *how*: Evidence from free relatives". In: *Linguistic Inquiry* 40, pp. 155–164.

Dayal, Veneeta (1996). *Locality in Wh Quantification: Questions and Relative Clauses in Hindi*. Dordrecht: Kluwer.

Groenendijk, Jeroen and Martin Stokhof (1982). "Semantic analysis of *wh*-complements". In: *Linguistics and Philosophy* 5, pp. 175–233.

Nicolae, Andreea C. (2020). "Functional multiple *wh* relatives". Talk presented at *NELS 51*.

Rudin, Catherine (1986). *Aspects of Bulgarian Syntax: Complementizers and WH Constructions*. Slavica Publishers.

Rudin, Catherine (2007). "Multiple *wh*-relatives in Slavic". In: *Formal approaches to Slavic linguistics*, pp. 282–307.

Rudin, Catherine (2008). "Pair-List vs. Single Pair Readings in Multiple *Wh* Free Relatives and Correlatives". In: *Kansas Working Papers in Linguistics 30*. The University of Kansas, pp. 257–267.

Xiang, Yimei (2021). "A hybrid categorial approach to question composition". In: *Linguistics and Philosophy* 44, pp. 587–647.

Future-past asymmetries, evidential grounding, and projection *

Fabrizio Cariani

University of Maryland, College Park College Park, MD, USA
fabrizio@umd.edu

1 Introduction

I develop a lexical solution to some important puzzles recently discovered by Ninan (forthcoming), which highlight striking asymmetries between future- and past-directed talk.

1.1 The asymmetry

Suppose at 6 PM I am cooking dinner for my kids before going out. Based on my extensive knowledge of their taste, I assertively utter *the kids will like this pasta*. In this context, my assertion seems flawless. Later, I receive a text message from the babysitter that says *the kids ate their dinner and are in bed now* (assume that this conveys no new information about whether the kids liked the pasta or not). Suppose someone asks: "Did the kids like their dinner?". In this context, the response *the kids liked the pasta* would be defective, for it would suggest I have evidence that stems directly from the event itself. Summarizing the pattern:

(1) a. **(early)** The kids will like this pasta. [good on the basis of the evidence]
b. **(late)** The kids liked this pasta. [bad on the basis of same evidence]

The puzzle is how to account for these asymmetries between future- and past-directed speech, while accepting that, in the early context, (1-a) is roughly equivalent to (1-b) in the late context.

1.2 My proposal in a nutshell

I develop the idea that the puzzle arises from the different evidential requirements imposed by modal items such as *will* as opposed to (tensed) bare predication. A view along these lines is sketched in chapter 13 of Cariani (2021) but that development is self-avowedly preliminary, especially with regards to how these requirements are to project under embeddings .

More specifically, I suggest that (1-b) is defective because ordinary predications (such as *the kids liked this pasta*) must be grounded in the following sense: the speaker's evidence for the relevant proposition must be causally upstream from the eventualities that the proposition is about. The reason why (1-a) is good, as uttered in its context, is that modals can modulate (i.e., either completely defuse, or reorient) these lexical requirements.

At the heart of my approach is a generalization of Willer & Kennedy's (2022) idea of *grounding*, understood as a "sincerity condition" on assertion—a lexically encoded requirement to the effect that the speaker's mental state must relate appropriately to the subject matter

*Thanks to the Meaning Group at the UMD, to the Konstanz Workshop on *Acquaintance Inference: Linguistic and Philosophical Perspectives*, to the 2022 Dubrovnik Conference on Philosophy of Language and to the Berlin Workshop in Philosophy of Language. Among the people in these audiences, special thanks to Pranav Anand, Sam Cumming, Daniel Drucker, Natasha Korotkova, Dilip Ninan, Paolo Santorio, and Malte Willer.

of the assertion. Willer & Kennedy go on to apply this idea to some specific puzzles, such as explaining how the acquaintance inference arises. In their analysis, adjectives like *tasty* encode acquaintance requirements—constraints to the effect that the speaker has had the relevant experience. These constraints percolate up through the semantic composition, so that *this apple is tasty* is grounded iff the speaker is acquainted with the specific apple.

While I borrow Willer & Kennedy's concept of grounding, my intended application requires that we take grounding constraints to apply well beyond the boundaries of subjective language. Furthermore, since I am only focused on cases in which grounding requires the availability of a specific kind of evidence, I will refer to this as *evidential grounding* (so as to keep the concept more clearly distinct from the concept of grounding that is prominent in metaphysics).

2 Quick overview of other positions

The nascent literature on Ninan's asymmetry puzzle has identified several alternative positions worth acknowledging. While there is no space here for a full critique—my focus being on positive development, rather than on shutting down alternatives—it is valuable to touch briefly on some of these arguments.

Ninan (forthcoming) critiques lexical views that impute the asymmetry to the hypothesis that *will* is a (non-veridical) necessity operator. Ninan leverages, among other things, arguments by Cariani and Santorio (2018) to the effect that *will* is not a necessity operator at all.

Ninan also refutes the idea that the asymmetry is due to differences in what evidence is *potentially available* to the speaker at the two times. After all, it is easy to modify the case so that the same kinds of evidence are available at the two times. For example, we can imagine that, before going to bed, both the kids and the babysitter take a pill which erases all their memories of the evening. Despite the fact that in this variant the earlier and later contexts share the same evidence, the judgments that drive the case do not substantively change.

Ninan's own positive account, which involves positing differences in how potential abnormalities can affect future- and past-directed knowledge is, in turn, critiqued in Cariani (2021), based on the observation that it is possible to imagine future-directed claims that are undermined by abnormalities in the past.

Another, more general, objection for any knowledge-based view adapts a version of an objection raised, in a different context, by Ninan himself (2014).[1] It seems plausible that we are in a position to assert (2-a) when talking about an apple we have not tried. Under the knowledge norm of assertion, if someone is in a position to assert (2-a), they must know its content. However, if *must* is veridical, (2-a) plausibly entails (2-b). If knowledge is closed under single-premise entailment, it follows that one must also know (2-b).

(2) a. This apple must be tasty.
b. This apple is tasty.

Similarly, in the case of future-past asymmetries, one is plausibly in a position to assert (3-a). If the knowledge norm of assertion is operative, one must know its content. If *must* is veridical and knowledge is closed under single-premise entailment, one must thus also know (3-b).

(3) a. The kids must have liked the pasta.
b. The kids liked the pasta.

[1] I am referring here to the discussion in Ninan (2014, p. 304-305). In Cariani (2021), I misattributed this argument, having forgotten that a variant of it had first shown up in the context of Ninan's own discussion of the acquaintance inference. I'll take this as a chance to correct the record.

Upshot: the unassertibility of (1-b)-(3-b) is not due to lack of knowledge, and so, the past-tense asymmetries are not asymmetries of knowledge. Of course, this argument is not a decisive refutation of knowledge-based views, since it can be blocked by denying any of (a) the knowledge norm of assertion; (b) the veridicality of *must*, or (c) the closure of knowledge under single-premise entailment. More argument is needed, but this is surely one direction in which that argument might develop.

3 General Structure of a Lexical Analysis

3.1 Grounding vs. directness.

The present lexical analysis claims that (1-b) demands a distinctive kind of evidence, whereas (1-a) does not. It is tempting to say that (1-b) demands *direct* evidence.(I myself did so in chapter 13 of Cariani 2021.) But there are reasons to avoid this conceptualization. The concept of directness is already recruited in the literature on the evidentiality of *must* and specifically in the literature on Kartunnen's problem (e.g. by von Fintel & Gillies 2010). There is no reason to think that the relevant evidential requirement must be the same across the two puzzles. If anything, the evidential contribution of *will* is known to be distinct from that of *must* (Winans 2016).

I propose that (1-b), and bare (present- and past-directed) predication generally, demands that the evidence be **grounded** in the event described by the prejacent, and relatedly that we describe certain lexical items as having **evidential grounding requirements**.

The exact analysis of grounded evidence can be investigated, in a modular fashion, as independent from the semantic questions that are focused on in this paper. My rough proposal is that we think of these grounding constraints in causal terms. Roughly the requirement is that the speaker's belief must be causally downstream from the eventualities that it is about.[2] For example, my belief that the kids enjoyed the pasta ought to be causally downstream from the event of their enjoying it.

Thinking of grounding in causal terms makes a remarkable prediction about scenarios in which the existence of known common causes supplies excellent inductive evidence for certain propositions, and yet seems insufficient to assert them. For example, suppose that, for whatever reason, we are trying to investigate whether the Post printed that some candidate for office, Ms. Brown let's say, won an election. I am unable to find a copy of the Post, but I can find a copy of the Times. The Times indeed says that Brown won the election. Naturally, there is a very strong correlation between what the Times and the Post decide to print. However, it is clear that I am not in a position to assert:

(4) The Post printed that Brown won the election.

3.2 Grounding constraints vs. ordinary presuppositions.

If grounding requirements are lexically encoded, they are plausibly not part of at-issue content, nor are they entailed by it. One way to see this, is to notice that *the kids did not like the pasta* is as much in need of grounding as (1-b), but *the kids must have liked the pasta* is not.[3]

Reflecting on how grounding constraints are to be treated as elements of meaning motivates a series of important desiderata:

[2] This was also the proposal of Cariani 2021, alongside the now-deprecated talk of *directness*.
[3] See, however, Dinges & Zakkou (2021), for a defense of the entailment view.

Desideratum 1. A model of not-at-issue content that is capable of capturing evidential grounding must be unlike the partial function model of presupposition.

After all, (1-b) does not go truth-valueless if the speaker's belief is not evidentially grounded.

Unlike ordinary presuppositions, grounding constraints do not project in conditional antecedents, questions, restrictors, and modals. Instead, their projection pattern appears closer to the projection pattern of acquaintance requirements (Ninan 2014, 2020, Anand and Korotkova 2018, Kennedy and Willer 2018, 2022). Here is a comparison of the two paradigms. The projection pattern of acquaintance requirements is in (5) and the analogous pattern for future-past asymmetries is in (6). I mark examples as '$\hookrightarrow \triangledown$' or $\not\hookrightarrow \triangledown$ depending on whether they involve a grounding requirement.

(5) a. That apple is tasty $\hookrightarrow \triangledown$
b. If that apple is tasty, he will purchase a whole bag of them. $\not\hookrightarrow \triangledown$
c. Is that apple tasty? $\not\hookrightarrow \triangledown$
d. The apple might be tasty. $\not\hookrightarrow \triangledown$
e. The apple will be tasty. $\not\hookrightarrow \triangledown$

(6) a. The kids did not like the pasta $\hookrightarrow \triangledown$
b. If the kids liked the pasta, they went to bed happy. $\not\hookrightarrow \triangledown$
c. Did the kids like the pasta? $\not\hookrightarrow \triangledown$
d. The kids might have liked the pasta. $\not\hookrightarrow \triangledown$
e. The kids will like the pasta. $\not\hookrightarrow \triangledown$

This pattern motivates:

Desideratum 2. Basic projection and obviation facts must be accounted for.

Failure of grounding requirements to project under questions is strongly suggestive that these requirements are not presuppositional. So this is plausibly viewed as a theory of projection that is additional to, and developed alongside, the standard theory of presupposition projection.

4 Disjunction and conjunctions.

An important challenge for any approach in this style comes from projection under disjunction and conjunction. Here are some examples and the desiderata they motivate:

(7) a. Either the kids liked the pasta or they liked the salad. $\hookrightarrow \triangledown$
b. Either the kids liked the pasta or they didn't. $\not\hookrightarrow \triangledown$
c. The kids liked the pasta and the salad. $\hookrightarrow \triangledown$
d. The kids did not both like the pasta and the salad. $\hookrightarrow \triangledown$

Some specific commentary on these examples is in order. I mark (7-a) as having a grounding requirement, because it generates a contrast with its future-directed counterpart (*either the kids will love the pasta or the kids will love the salad*). This requirement is subtler, however, than the grounding requirement of the bare (1-a). One way to appreciate this, is that there are disjunctions like (7-b) that have no grounding requirements at all. Another crucial observation is that (7-c) and (7-d) do not have the same grounding requirements. For example, one can assert (7-d) merely on the basis of grounded evidence that the kids did not like the pasta.

Summing up, these examples motivate four more desiderata (using 'GR(A)' to abbreviate ⌜the grounding requirement of A⌝):

Desideratum 3. Substantive GRs must project through disjunction (GR(A or B) $\neq \top$).

Desideratum 4. GR(A or B) is less stringent than GR(A)$\cup$GR(B).

Desideratum 5. In light of (7-c), GR(A and B) = GR(A)$\cap$GR(B)

Desideratum 6. In light of (7-d), GR($\neg$(A and B)) $\neq$ GR(A and B).

Ultimately, **Desideratum 6** will require the projection system to be bilateral.

5 The Formal System

I propose a formal system for modeling grounding requirements and their projection behavior that meets all of these desiderata. To account for **Desideratum 1**, the system deploys a multi-dimensional model of meaning (parallel to treatments of presupposition in Kartunnen & Peters 1979, Sudo 2012, Mandelkern 2016). In this model, truth-conditions and grounding requirements may hold, or fail to hold, independently. For example, an utterance of (1-b), expresses a proposition that is fully evaluable for truth or falsehood, even if one of its grounding constraints is false. In fact, the present theory even entails that such propositions may be *known*.

The combination of desiderata outlined in the previous section motivates some additional design principles. In particular, as noted, **Desideratum 6** motivates a bilateral system: in a unilateral projection system A and its negation would have to share their *GR*s, which contradicts the Desideratum. Furthermore, the combination of **Desiderata 3** and **4** requires attention to how the grounding requirements may be satisfied, beyond satisfaction of the GRs ot the individual disjunct. My hypothesis is that (7-b) shows that arguments from the negation of one conjunct to the second conjunct may also clear the grounding requirements of a disjunction.

Vector-based meanings

Moving on to the specifics of the system, I model meanings as n-tuples, where the first coordinate is the contribution to truth-conditions, the second coordinate is the 'positive grounding' (=the grounding requirement in a positive context), and the third coordinate is the 'negative grounding' (=the grounding requirement that would emerge in negative environment, e.g. if negated). I omit ordinary presuppositions to simplify notation: they would be assigned another coordinate. Given expression e, let $[e]$ be the meaning vector of e, and $[e]_n$ its n-*th* coordinate.

Next, stipulate:

> **Core grounding principle** An assertive utterance of A is defective unless the positive grounding requirement, $[\mathsf{A}]_2$, is satisfied in the context of utterance.

Because in the present setting, grounding requirements are modeled as possible world propositions, the satisfaction of A's grounding requirements amounts to the claim that the second coordinate of A's meaning vector is true at the world of the context.

Entries for the sentential connectives

The entries below regulate the behavior of the sentential connectives. Let $g(\mathsf{A}, \mathsf{B})$ stand for the proposition that is true in world x if there is an argument available in context from A to B.

(8) a. **Negation:** $[not\ \mathsf{A}] = \langle W - [\mathsf{A}]_1, [\mathsf{A}]_3, [\mathsf{A}]_2 \rangle$
b. **Disjunction:** $[\mathsf{A}\ or\ \mathsf{B}] = \langle [\mathsf{A}]_1 \cup [\mathsf{B}]_1,\ [\mathsf{A}]_2 \cup [\mathsf{B}]_2 \cup g(\neg\mathsf{A}, \mathsf{B}) \cup g(\neg\mathsf{B}, \mathsf{A}),\ [\mathsf{A}]_3 \cap [\mathsf{B}]_3 \rangle$.

c. **Conjunction:** $[\mathsf{A}\ \textit{and}\ \mathsf{B}] = \langle [\mathsf{A}]_1 \cap [\mathsf{B}]_1,\ [\mathsf{A}]_2 \cap [\mathsf{B}]_2,\ [\mathsf{A}]_3 \cup [\mathsf{B}]_3 \cup g(\mathsf{A}, \neg\mathsf{B}) \cup g(\mathsf{B}, \neg\mathsf{A})\rangle$.

For each connective, the first coordinate merely registers the standard Boolean behavior. Note also that our theory about negation is a bit subtler than the original hypothesis in which it fixes grounding requirements. Instead, it swaps positive and negative requirements.

Entries for obviators

Our framework allows a particularly simple treatment for the phenomenon of obviation. Obviators are those expressions that disable grounding requirements. In the present system, they do this by intervening on the grounding coordinates of their prejacent's meanings. Consider, for instance, the modals *might* and *will*. For the sake of determinacy and simplicity, give *might* a contextualist semantics where it quantifies over a contextually determined modal domain $d_{c,w}$ and *will* a selection semantics (in the style of Cariani & Santorio 2018, Cariani 2021). Let $\top$ stand for a tautology. Sample vector meanings for these items might be:

(9) a. **Might**: $[\textit{might}\ \mathsf{A}]^{c,w} = \langle \lambda w. \exists v \in d_{c,w}, [\mathsf{A}]_1^{c,v}, \top, \top\rangle$
b. **Will**: $[\textit{will}\ \mathsf{A}]^{c,w} = \langle \lambda w. [\mathsf{A}]_1^{c,sel(h_c(w),w)}, \top, \top\rangle$

In both cases, the modals obviate by wiping the evidential constraints that derive from their prejacents, and replacing them with the tautological constraint.

6 Illustrations

Let s_c be the speaker in context c and let's work with the simplest of our examples. To derive the fact that *Ms. Brown won* and *Ms. Brown did not win* have the same grounding requirement, suppose that at the level of the semantics of the verb things start out symmetric.

(10) $[\text{won}]^{c,w} = \langle \lambda x \lambda w.\ x$ won in w,
$\lambda x \lambda w.\ s_c$ has downstream evidence in w that settles whether x won,
$\lambda x \lambda w.\ s_c$ has downstream evidence in w that settles whether x won$\rangle$

To derive the full semantics of *Ms. Brown won*, compose this point-wise with $[Ms.Brown]$ i.e., $\langle Ms.Brown, Ms.Brown, Ms.Brown\rangle$.

(11) [Ms. Brown won]$^{c,w} = \langle \lambda w.$ Ms. Brown won in w,
$\lambda w.\ s_c$ has downstream evidence in w that settles whether Ms. Brown won,
$\lambda w. s_c$ has downstream evidence in w that settles whether Ms. Brown won$\rangle$

Because the positive and negative side are identical, *Ms. Brown did not win* differs from *Ms. Brown won* only in its at-issue coordinate.

Finally, the bilateral system's power shows up for conjoined and disjoined grounding requirements the system, as shown by the contrast between the semantic values A and B versus that of $\neg(A$ and $B)$.

(12) $[\mathsf{A}\ \text{and}\ \mathsf{B}] = \langle [\mathsf{A}]_1 \cap [\mathsf{B}]_1,\ [\mathsf{A}]_2 \cap [\mathsf{B}]_2,\ [\mathsf{A}]_3 \cup [\mathsf{B}]_3 \cup g(\mathsf{A}, \neg\mathsf{B}) \cup g(\mathsf{B}, \neg\mathsf{A})\rangle$.

(13) $[\neg(\mathsf{A}\ \text{and}\ \mathsf{B})] = \langle \overline{[\mathsf{A}]_1 \cap [\mathsf{B}]_1}, [\mathsf{A}]_3 \cup [\mathsf{B}]_3 \cup g(\mathsf{A}, \neg\mathsf{B}) \cup g(\mathsf{B}, \neg\mathsf{A})\rangle, [\mathsf{A}]_2 \cap [\mathsf{B}]_2\rangle$

Because the two sentences differ in their positive GR, their acceptability conditions differ not just with respect to the at-issue content, but also with respect to their GR.

References

Anand, Pranav and Korotkova, Natasha. 2018. Acquaintance content and obviation, *Proceedings of SuB* 2018.

Cariani, Fabrizio. 2021. *The Modal Future.* Cambridge University Press.

Cariani, Fabrizio and Santorio, Paolo. 2018. *Will* done better. *Mind* 505(127). 129-165.

Dinges, Alexander and Zakkou, Julia. Taste, traits, and tendencies. *Philosophical Studies* 178(4). 1183-1206

von Fintel, Kai & Gillies, Thony. 2010. Must... Stay... Strong. *Natural Language Semantics* 18, 351–383.

Karttunen, Lauri & Stanley Peters. 1979. Conventional implicature. *Syntax and Semantics* 11. 1–56.

Kennedy, Christopher and Willer, Malte. 2018. Subjective attitudes and counterstance contingency. In *Proceedings of SALT 26*, 913-933.

Mandelkern, Matthew. 2016. Dissatisfaction theory. In *Proceedings of SALT 26*, 391–416.

Ninan, Dilip. 2014. Taste Predicates and the Acquaintance Inference. In *Proceedings of SALT 24*, 290-309.

Ninan, Dilip. 2020. The Projection Problem for Predicates of Taste. In *Proceedings of SALT 30*, 753-778.

Ninan, Dilip. (forthcoming). Assertion, Evidence, and the Future. *Philosophical Review.*

Sudo, Yasutada. 2012. On the semantics of Phi features on pronouns: *MIT PhD dissertation.*

Willer, Malte and Kennedy, Christopher. 2022. Assertion, expression, and experience. *Inquiry: An Interdisciplinary Journal of Philosophy* 65(7). 821-857.

Winans, Lauren. 2016. Inferences of *will*: *UCLA PhD dissertation.*

Disambiguating *will*-conditionals*

Ivano Ciardelli[1] and Adrian Ommundsen[2]

[1] Università di Padova, ivano.ciardelli@unipd.it
[2] Ludwig-Maximilians Universität München, adriankommundsen@gmail.com

Abstract

We bring together two ideas that are widely accepted in the literature: (i) conditionals with modal consequents are systematically ambiguous between a reading on which the if-clause restricts the modal, and a reading on which the if-clause introduces an epistemic supposition and the modal is unrestricted; (ii) 'will' is a modal. Together, (i) and (ii) imply that 'will'-conditionals are systematically ambiguous between a reading as restricted historical claims, and one as unrestricted historical claims made under a supposition. We argue that this prediction is correct and that the existence of the relevant ambiguity has far-reaching repercussions for several important debates concerning conditionals.

1 Introduction

In this paper we discuss how two widely accepted ideas, when taken together, have important implications for the analysis of *will*-conditionals. The first idea is that conditionals with modals in consequent are ambiguous in a systematic way [12]: on one reading, called the *O-reading*, they are restricted modal claims; on another reading, called the *C-reading*, they are conditional assertions of unrestricted modal claims. The second idea is that *will* is a modal: like other modals, its semantics involves a modal base, which may be restricted by *if*-clauses [1, 4, 5, 18]. Together, these ideas imply that conditionals of the form "if A, will B" are systematically ambiguous between an O-reading on which they express a restricted historical claim, and a C-reading on which they are conditional assertions of unrestricted historical claims.

In this paper we have two related goals. The first goal is to provide empirical evidence in support of the existence of the C/O ambiguity for *will*-conditionals. The second goal is to show that the existence of this ambiguity has important implications for several long-standing debates about conditionals, including the debate on the relation between past tense indicatives and counterfactuals, the debate on whether *will*-conditionals belong with the former or with the latter kind, and the debate on the relations between conditionals and supposition.

To be clear, we are by no means the first to claim that *will*-conditionals are ambiguous (see in particular [24, 19]). However, our aim is not just to show that such conditionals are ambiguous, but to show that they are ambiguous in the particular way predicted by a plausible theoretical story. Moreover, we believe the far-reaching repercussions of the existence of this ambiguity have not yet been sufficiently appreciated, and we hope to make a contribution in this direction.

2 Background: The C/O ambiguity

Geurts [12] has pointed out that conditionals whose consequent contains a modal or an adverbial quantifier are systematically ambiguous. For an illustration, consider (1):

*Thanks to David Boylan, Fabrizio Cariani, Magda Kaufmann, Matt Mandelkern, Paolo Santorio, Una Stojnic, and three AC reviewers for precious feedback on the ideas presented here and pointers to the literature.

(1) If Beryl is in Paris, she often visits the Louvre.

On one reading, (1) is a claim of conditional frequency: on many of the occasions on which she's in Paris, Beryl visits the Louvre. As this paraphrase makes clear, this reading arises when the *if*-clause restricts the set of occasions that the adverbial quantifier *often* ranges over. Geurts calls this the *O-reading* of (1), since the *if*-clause restricts the range of the Overt operator *often*.

On another reading, (1) is a conditional claim of unconditional frequency: it asserts that, under the supposition that Beryl is in Paris, we can conclude that she visits the Louvre often. (Imagine that Beryl is an art lover who is spending her sabbatical in a European capital, but we don't know which one.) In this case, the role of the *if*-clause is not to restrict the range of the quantifier *often*, but rather to introduce a supposition, just as in the typical examples of indicative conditionals. This effect is naturally modeled by taking the *if*-clause to restrict a set of epistemic possibilities, so that the consequent is asserted relative to this restricted set. Geurts calls this the *C-reading* of (1) since, following Kratzer [21], he takes it to arise from the *if*-clause restricting a Covert epistemic modal (as discussed in [7], this assumption is not unavoidable; however, the details of how the C-reading arises are not crucial for our purposes).

For an illustration of how the C/O ambiguity arises with modals, suppose Alice and Bob are colleagues and consider:

(2) If Alice leaves, Bob must stay.

On one reading, (2) is a statement of conditional obligation: it is true in a situation in which one of Alice and Bob has to be present—it doesn't matter who. In this situation, Bob is not required to stay, but he is required to stay *conditionally on Alice leaving*. This arises naturally as an O-reading, with the *if*-clause acting as a restrictor of the deontic modal base [20]: on this analysis, (2) asserts that all permissible worlds where Alice leaves are worlds where Bob stays.

On another reading, (2) is a conditional claim of unrestricted obligation: this reading is true in a situation in which a specific employee—Alice or Bob, we don't know who—is required to be present. Given Alices's conscientiousness, if she leaves we may conclude that it is Bob who is required to stay. This is naturally analyzed as a C-reading, with the antecedent introducing an epistemic supposition and the deontic modal remaining unrestricted.

It will be useful to fix a notation that allows us to refer quickly to the LFs that give rise to the two readings in question. In case case of (2), letting A and B stand, respectively, for "Alice leaves" and "Bob stays", we will denote the relevant LFs as follows:

$$\text{O-reading: } \mathsf{must}_{\mathsf{A}}(\mathsf{B}) \qquad \text{C-reading: } \mathsf{A} \rightarrow \mathsf{must}(\mathsf{B})$$

One should not read too much into this notation: in particular, we do not mean to suggest that conditionals are lexically ambiguous between a restrictor reading and an epistemic reading. On the contrary, we are convinced that a uniform analysis in terms of restriction behavior is right. But there are different proposals for such an analysis (e.g., see [12] and [7]), and we need not settle on one here. The above notation has the merit of being suggestive of the crucial point, namely, whether or not the antecedent restricts the modal in the given LF.

As Geurts [12] has argued, C/O ambiguity arises generally for conditionals involving modals in the consequent. In recent years, several authors have argued that *will* is a modal and have proposed accounts of *will* within the general framework of modal semantics [1, 4, 5, 18]. We thus expect the C/O ambiguity to arise, in particular, for conditionals whose consequent contains *will*. Since these conditionals play such a prominent role in the way we think and talk about the future, formulate predictions, and make decisions, it seems crucial to ask whether this expectation is borne out. In the next section, we will argue that it is.

3 The C/O ambiguity in *will*-conditionals

Consider Gibbard's [13] famous riverboat scenario. Sly Pete is playing poker with Mr. Stone on a Mississippi riverboat. Stone has bet up to the limit for the hand, and it is now up to Pete to call or fold. Secretly, Zack has signalled to Pete the content of Stone's hand. Now consider:

(3) If Pete calls, he will win.

Is (3) true? Intuitively, there seem to be two natural answers.

- Answer 1: uncertain. If Pete has a winning hand, then (3) is true; but if he has a losing hand, (3) is false. We don't know whether Pete has a winning hand or a losing hand, and so we don't know whether (3) is true.[1]
- Answer 2: true. Pete knows his opponent's cards, is a competent poker player, and plays to win: he is not going to call with a losing hand. So if he calls, that means he has a winning hand, and so he will win. Therefore, (3) is true.

These two reactions are exactly what we would expect to see if (3) allows both for an O-reading and for a C-reading. Let us see why.

Consider first the C-reading, $\mathsf{call} \rightarrow \mathsf{will}(\mathsf{win})$. To assess it, we have to consider what we can infer from the epistemic supposition that Pete calls. Given the context, one thing we can infer is that he has winning cards, and then, from this and the supposition that he calls, we can infer that he will win. Hence, this reading of (3) is true. To spell out this prediction a bit more formally, suppose an epistemic conditional $\mathsf{A} \rightarrow \mathsf{B}$ is true if in every epistemic possibility where A is true, B is true.[2] In the described scenario, we may take ourselves to know that Pete will call only if he has a winning hand. So, our information state rules out worlds where Pete calls with a losing hand. As a consequence, every world in our epistemic state where Pete calls is a world where he calls with a winning hand and, therefore, a world where he will win. So, in every epistemic possibility where the antecedent of $\mathsf{call} \rightarrow \mathsf{will}(\mathsf{win})$ is true, the consequent is true. As a consequence, the C-reading of (3) is predicted to be true, which explains Answer 2.

Let us now turn to the O-reading, $\mathsf{will}_{\mathsf{call}}(\mathsf{win})$. In order to see what is predicted in this case, we need to build on a theory of *will*. We adopt the theory of Cariani and Santorio [4, 5] (henceforth: C&S), although other theories (for instance, the one by Kaufmann in [18]) would lead to the same predictions in our example. According to C&S's theory, the semantics of *will* involves two parameters: a modal base m, which is initialized by context to the map sending a world w to the set $m(w)$ of worlds which are duplicates of w up to the time of utterance; and a selection function f, which maps a world w and a set of worlds X to a world $f(w, X)$ with the constraints that (i) $f(w, X) \in X$ whenever $X \neq \emptyset$, and (ii) $f(w, X) = w$ whenever $w \in X$. A claim $\mathsf{will}(\mathsf{A})$ is true at a world w relative to m and f in case A is true at $f(w, m(w))$.

Now consider a world w in which Pete has a losing hand. All worlds in the historical modal base $m(w)$ (i.e., all historical duplicates of w up to the time of utterance) are worlds where Pete has a losing hand. Now suppose we restrict $m(w)$ to worlds where Pete calls. Then every world in the restricted set is a world where Pete calls with a losing hand and therefore loses. So no matter which world is picked by the selection function for will, the prejacent win is false at that world.Therefore, the restricted modal claim $\mathsf{will}_{\mathsf{call}}(\mathsf{win})$ is false at w. Reasoning in a

[1] Another way to elicit this intuition is to ask ourselves: should Pete call? Well, not necessarily. After all, it is not obvious that if he calls he will win: it depends on the cards he has.

[2] On several accounts (e.g., [6, 14, 27, 30, 31]) epistemic conditionals are not strictly speaking true or false. On these accounts, our informal talk of a conditional being *true* just tracks the fact that the conditional is *fully acceptable*, or *supported*, on the basis of the availabe information. This does not affect our argument here.

similar way, we can see that in a world w where Pete has a winning hand, the claim $\mathsf{will}_{\mathsf{call}}(\mathsf{win})$ is instead true. So, the prediction is that on the O-reading, (3) is true if and only if Pete has winning cards. Since we are uncertain whether Pete has winning card, we are uncertain about whether the O-reading is true. This accounts for Answer 1 above.

Summing up: empirically, there seem to be two natural ways to react to (3) in the context we described. These two reactions are exactly what we would expect to see if (3) is ambiguous between a C-reading and an O-reading. So, there is empirical reason for thinking that *will*-conditionals are not an exception: just like other modal conditionals, they are ambiguous in a systematic way, and this ambiguity must be taken into account when theorizing about them. In the rest of this paper we will see how taking into account this ambiguity has important repercussions for some long-standing debates about conditionals.

4 The classification of conditionals

The literature is dominated by a classification of conditionals into two kinds: *indicatives*, and *subjunctives*. The distinction is motivated by the following contrast, pointed out by Adams [2]:

(4) a. If Oswald didn't kill Kennedy, someone else did.
b. If Oswald hadn't killed Kennedy, someone else would have.

The two sentences are judged and interpreted differently. (4-a) receives an epistemic reading: in supposing the antecedent, we revise our beliefs about what the world is like; (4-b) receives a historical reading: in supposing the antecedent, we entertain a non-actual course of history.

It has been supposed that the difference is due to the morphology of the relevant antecedents, and many proposals have been formulated [9, 28, 29, 30] to account for the difference in terms of the contribution of indicative or subjunctive marking.[3] But, is the basic assumption correct?

We saw that the difference between epistemic readings and historical readings arises even for a single *will*-conditional like (3). To see that this is the same difference manifested by the pair in (4), take again the riverboat scenario, and suppose we leave the room before Pete calls or folds. Time goes by: the events are now in the past, but our information has not changed. Now we are clearly in a position to accept (5-a): if Pete called, that means he had a winning hand, so he won. But we are not in a position to accept (5-b): after all, perhaps Pete had a losing hand and folded; in that case, if he had called, he would have lost; so (5-b) may be false.

(5) a. If Pete called, he won.
b. If Pete had called, he would have won.

This strongly suggests that (5-a) shares the acceptable epistemic reading of (3), while (5-b) shares the uncertain historical reading of (3). Hence, the difference between the past indicative (5-a) and the past *would*-conditional (5-b) presumably comes about in just the same way as the difference between the two readings of (3). And this seems independently plausible: after all, (5-a) contains no modal to restrict, so only an epistemic reading is available; by contrast, (5-b) contains the historical modal *would*, which may be restricted by the *if*-clause yielding a historical reading. The same goes, of course, for Adams' pair in (4). To be clear, we are not denying that mood and tense plays an important semantic role in conditionals. However, we believe the previous discussion indicates that the very different interpretation of the Adams' sentences in

[3] Or, perhaps more appropriately, the contribution of *O-marking* and *X-marking* [11]; it has been argued that it many languages, including English, the relevant distinction is not one of indicative vs. subjunctive mood.

(4) is due primarily to the fact that the *if*-clause targets different semantic parameters in the two cases: a set of epistemic possibilities in (4-a), a set of historical possibilities in (4-b).

This perspective also suggests an alternative to the traditional classification of conditionals into indicatives and subjunctives. Indeed, on the basis of the restrictor view, a semantically natural way to classify conditionals is according to the parameter targeted by the *if*-clause: we will thus have epistemic conditionals, historical conditionals, deontic conditionals, etc., depending on whether the *if*-clause restricts a set of epistemic, historical, or deontic possibilities. Of course, this classification applies at the level of LF; so for instance, (2) may be interpreted either as a deontic conditional or as an epistemic conditional with a deontic consequent; similarly, (3) may be interpreted as a historical conditional or as an epistemic conditional with a historical consequent. We will close the paper by illustrating the merits of this alternative classification.

5 The indicative constraint

It is widely held that indicative conditionals require their antecedents to be epistemically possible; if this condition is not met, an indicative conditional cannot be used felicitously. E.g., if we are certain that Pete didn't call, an utterance of (6) is infelicitous.

(6) #If Pete called, he lost.

However, there are apparent counterexamples to this claim. Some of these counterexamples involve *will*-conditionals. For instance, suppose I tell you that Pete will definitely not call. You accept this, but you want to know *why* he won't call. As an explanation, I offer:

(7) If he calls, he will lose. And he knows that.

My utterance seems perfectly felicitous. For another counterexample, this time involving an adverbial quantifier, consider the following discourse.

(8) A: How about inviting Beryl for dinner?
B: She's not in town.
A: Oh, I see. How do you know?
B: If she's in town, she always gives me a call.

Again, the conditional is perfectly felicitous. How should we make sense of these observations?

The solution is to avoid sweeping generalizations about indicative conditionals, and instead pay attention to the more fine-grained classification discussed in the previous section. Some indicative conditionals are epistemic, but some are not, and their semantic properties are bound to be different from those of epistemic conditionals. The correct generalization, in our view, is: conditionals require their *if*-clause to be true at some possibility *within the domain it targets*.[4]

Let us see how this proposal allows us to explain the data above. In the case of epistemic conditionals, the relevant domain of possibilities is the set of *epistemic* possibilities open in the context. So, for these conditionals our requirement boils down to the standard requirement that the antecedent be epistemically possible, explaining the infelicity of (6).

However, in the case of historical conditionals, the *if*-clause does not target a set of epistemic possibilities, but a set of *historical* possibilities—the set of histories open at a given time.

[4]This subsumes as special cases proposals that have been often made before about epistemic conditionals (see, e.g., [9, 15]) and counterfactuals (see, e.g., [10, 16]). Also, note that the requirement we propose amounts to the condition that the domain resulting from the restriction with an *if*-clause be nonempty; so, it is presumably grounded in a more general ban against empty domains of quantification.

The requirement will then be that the antecedent be true at some of these possibilities—i.e., historically possible. In (7), since Pete has not yet acted, the possibility of calling is presumably regarded as historically open by the speaker, which explains why the conditional is felicitous.[5]

Finally, in (8), the relevant interpretation of the conditional is one where the *if*-clause restricts a set of relevant occasions. Thus, the requirement amounts to the existence of occasions on which Beryl was in town. This requirement may well be satisfied even though Beryl is not in town on the current occasion, which explains why the conditional is felicitous.

6 Conditionals and supposition

The *Ramsey test* idea, in its general version, is the idea that to bear an attitude to a conditional is to bear that attitude to the consequent on the supposition of the antecedent. Thus, e.g., to accept a conditional is to accept the consequent on the supposition of the antecedent; to have a certain degree of credence in a conditional is to have that degree of credence in the consequent on the supposition of the antecedent (Adams' thesis); and so on. The idea has great psychological appeal and is well-supported by empirical data. Nevertheless, the literature contains apparent counterexamples [17, 22, 23, 25, 26]. Most of them (but not all) involve *will*-conditionals.[6]

To get a sense of how these counterexamples work, consider again our conditional (3) in the scenario above. We saw that, on one reading, this conditional is true if and only if Pete has a winning hand. So, the probability of (3) under this reading is just the probability that Pete has a winning hand; we may suppose that this probability is low, since (let us stipulate) we know that Stone has an excellent hand. At the same time, since we consider it nearly certain that Pete will call only if he has a winning hand, the conditional probability of Pete winning given that he in fact calls is very high. So, it would appear, we have low credence in the conditional (3), even though we have high credence in the consequent on the supposition of the antecedent, in contrast with Adams' thesis and with the Ramsey test idea.

But this conclusion is mistaken: what we are doing here is to compare our attitude towards a *historical* conditional, $\mathsf{will}_{\mathsf{call}}(\mathsf{win})$, with our attitude to the consequent under the *epistemic* supposition of the antecedent. However, the correct way to understand the Ramsey test is as relating kinds of conditionals to corresponding kinds of supposition (cf. [3]): our attitude to an *epistemic* conditional $\mathsf{A} \to \mathsf{B}$ should reflect our attitude to B under the epistemic supposition of A; but our attitude to a *historical* conditional $\mathsf{will}_{\mathsf{A}}(\mathsf{B})$ should reflect our attitude to B under the *historical* supposition of A. In particular, since Bayesian conditionalization corresponds to the *epistemic* mode of supposition, Adams' thesis that the probabilities of conditionals are conditional probabilities should be understood as a claim about the probabilities of *epistemic* conditionals. It is not expected to apply to historical readings of *will*-conditionals.

In this way, many (though not all) alleged counterexamples to the Ramsey test and to Adams' thesis can be diagnosed as stemming from equivocation. Once the ambiguity of *will*-conditionals is properly taken into account, these seeming counterexamples are dispelled.

[5] An interesing observation, due to Matt Mandelkern (p.c.) is that in contrast to (i-a), (i-b) is infelicitous:

(i) a. Pete won't call. If he calls, he will lose.
b. Pete won't call. #If he will call, he will lose.

We may explain this if we assume that the historical modal base for *will* may not be restricted by a *will*-clause (this requires an explanation, but does not seem implausible). This rules out a historical interpretation of the conditional in (i-b), leaving an epistemic interpretation as the only option, with the associated requirement that the antecedent be epistemically possible.

[6] Out of our list of references, the exception is McGee's counterexample [23], which does not use *will*-conditionals. See [8] for a possible response.

References

[1] Dorit Abusch. Sequence of tense and temporal de re. *Linguistics and philosophy*, pages 1–50, 1997.

[2] Ernest Adams. Subjunctive and indicative conditionals. *Foundations of Language*, 6(1):89–94, 1970.

[3] Richard Bradley. A defence of the ramsey test. *Mind*, 116(461):1–21, 2007.

[4] Fabrizio Cariani. *The modal future: A theory of future-directed thought and talk*. Cambridge University Press, 2021.

[5] Fabrizio Cariani and Paolo Santorio. Will done Better: Selection Semantics, Future Credence, and Indeterminacy. *Mind*, 127(505):129–165, 02 2018.

[6] Ivano Ciardelli. Restriction without quantification: embedding and probability for indicative conditionals. *Ergo*, 8(21):598–636, 2021.

[7] Ivano Ciardelli. The restrictor view, without covert modals. *Linguistics and Philosophy*, 45:1–28, 2022.

[8] Ivano Ciardelli and Adrian Ommundsen. Probabilities of conditionals: updating Adams. *Noûs*, 2022.

[9] Kai von Fintel. The presupposition of subjunctive conditionals. *The interpretive tract*, 25:29–44, 1998.

[10] Kai von Fintel. Counterfactuals in a dynamic context. *Current Studies in Linguistics Series*, 36:123–152, 2001.

[11] Kai von Fintel and Sabine Iatridou. Prolegomena to a theory of X-marking. Forthcoming in *Linguistics and Philosophy*.

[12] Bart Geurts. On an ambiguity in quantified conditionals. Manuscript, University of Nijmegen. Available at http://ncs.ruhosting.nl/bart/papers/conditionals.pdf, 2004.

[13] Allan Gibbard. Two recent theories of conditionals. In *Ifs: Conditionals, beliefs, decision, chance, time*, pages 211–247. Reidel, Dordrecht, 1980.

[14] Anthony Gillies. Epistemic conditionals and conditional epistemics. *Noûs*, 38(4):585–616, 2004.

[15] Anthony Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

[16] Anthony Gillies. On truth-conditions for if (but not quite only if). *Philosophical Review*, 118(3):325–349, 2009.

[17] Stefan Kaufmann. Conditioning against the grain. *Journal of Philosophical Logic*, 33(6):583–606, Dec 2004.

[18] Stefan Kaufmann. Conditional truth and future reference. *Journal of semantics*, 22(3):231–280, 2005.

[19] Justin Khoo. On indicative and subjunctive conditionals. *Philosophers' Imprint*, 15(32):1–40, 2015.

[20] Angelika Kratzer. The notional category of modality. In H.J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts: New Approaches in Word Semantics*, pages 38–74. Berlin: Walter de Gruyter, 1981.

[21] Angelika Kratzer. Conditionals. *Chicago Linguistics Society*, 22(2):1–15, 1986.

[22] Matthew Mandelkern and Justin Khoo. Against preservation. *Analysis*, 79(3):424–436, 2019.

[23] Vann McGee. To tell the truth about conditionals. *Analysis*, 60(265):107–111, 2000.

[24] Adam Morton. Indicative versus subjunctive in future conditionals. *Analysis*, 64(4):289–293, 2004.

[25] Sarah Moss. *Probabilistic Knowledge*. Oxford University Press, 2018.

[26] Daniel Rothschild. Do indicative conditionals express propositions? *Noûs*, 47(1):49–68, 2013.

[27] Paolo Santorio. Path semantics for indicative conditionals. *Mind*, 131(521):59–98, 01 2022.

[28] Katrin Schulz. Conditionals from a linguistic point of view: Two case studies. *Journal of Philosophical Logic*, 44(6):805–816, 2015.

[29] Robert Stalnaker. Indicative conditionals. In *Ifs*, pages 193–210. Reidel, 1976.

[30] William Starr. A uniform theory of conditionals. *Journal of Philosophical Logic*, 43(6):1019–1064, 2014.

[31] Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

A minimally sufficient analysis of sufficiency modal constructions *

Cleo Condoravdi[1] and Itamar Francez[2]

[1] Stanford University, Stanford, CA, U.S.A
cleoc@stanford.edu
[2] University of Chicago, Chicago, IL, U.S.A.
ifrancez@uchicago.edu

Abstract

A well-known compositionality problem involves the interaction between *only* and goal-oriented modality: *only* can weaken a necessity statement and a necessity modal can suspend a characteristic inference associated with *only*, the inference to the prejacent. Existing proposals resort to non-standard assumptions either about *only* or what it combines with. We show that a standard analysis of *only* and a plausible analysis of goal-oriented modality can account for the compositionality problem.

1 The prejacent problem

The *sufficiency modal construction*, exemplified in (1), gives rise to a puzzle that von Fintel & Iatridou [16] (vF&I) dubbed "the prejacent problem". This problem involves the compositional interaction between *only* and goal-oriented (aka teleological) modality, and can be appreciated by comparing (1) with (2), where (2) corresponds to the prejacent of *only* in (1).

(1) To get good chocolate, you only have to go [to Trader Joe's]$_F$.

(2) To get good chocolate, you have to go to Trader Joe's.

Only appears to weaken the necessity statement in its scope: while (2) implies that there is only one way to get good chocolate, (1) is compatible with there being other ways, but implies that those ways are more involved, in some sense, than going to Trader Joe's. A necessity modal under *only*, in turn, can suspend the inference to the prejacent that is characteristic of *only*: while (3) commits the speaker to (4), (1) does not commit the speaker to (2).

(3) You only have to go [to Trader Joe's]$_F$ today.

(4) You have to go to Trader Joe's today.

This inferential pattern thus provides a challenge to a uniform analysis of *only* and raises the question of what its presuppositional content is.

An adequate analysis of sufficiency modal constructions should satisfy certain desiderata. The analysis should determine an interpretation for *only* in the construction and say how that interpretation relates to its interpretation elsewhere. Ideally, *only* would receive the same interpretation everywhere. It must also capture two implications of (1) One, which we refer to as the *actual means implication*, is that going to Trader Joe's is actually a way of getting chocolate. The other, which we refer to as the *scalar implication*, is the implication, pointed

*We thank audiences at a UConn colloquium, the CSLI Workshop on Logic, Rationality, and Intelligent Interaction, the Workshop on Alternatives, Expectations and Domain Widening at Bar Ilan University, the Construction of Meaning workshop at Stanford, and three anonymous reviewers for comments and discussion.

out by Krasikova & Zechev [12] as part of their critique of vF&I's non-scalar analysis, that lower ranked alternatives to the modal prejacent, such as going to a store that is closer, are *not* actual means to this goal. If your corner store is closer than Trade Joe's and both have good chocolate, (1) cannot be truly and felicitously uttered; instead, (5) has to be used.

(5) To get good chocolate, you only have to go to your corner store.

None of the existing analyses of sufficiency modal constructions [16, 10, 12, 8, 11, 1, 2] fully meet these desiderata. Those that build scalarity into the meaning of the construction nevertheless resort to non-standard and, we argue, analytically problematic assumptions either about the compositional make-up of *only* and/or its prejacent, or about *only*'s presuppositional content. This paper demonstrates that a preference-based analysis of goal-oriented modals, coupled with a standard scalar analysis of *only*, provides a compositional account that derives the interpretive behavior of the sufficiency modal construction.

2 Analysis

Descriptively, the sufficiency modal construction consists of *only* and a necessity clause composed of a modalized sentence with a necessity modal modified by a purpose clause. The analytical question is whether these two components suffice for a compositional analysis. On the face of it, the existence of the prejacent problem indicates that the compositional structure is more complex. Complicating the compositional structure has been the approach taken by all previous accounts except for Franke's [8]. We pursue an analysis in which the compositional structure of the construction reflects its grammatical structure. Our main claim is that the interpretive properties of the construction are due (a) to the contextual alternatives activated by *only* and (b) the contextual resolution of the parameters of the modal in the prejacent of *only*. We first discuss the interpretation of goal-oriented modality, then the interpretation of *only* and the effect it has on the interpretation of the necessity clause in its scope.

2.1 Goal-oriented modality

Consider the exchange in (6), where a goal-oriented necessity clause is used in response to an assertion about a search for good chocolate. Intuitively, B's response in (6) provides the factual information that Trader Joe's carries good chocolate and that it's the only (relevant) shop that does. More directly tied to its semantic content, it is used by B to communicate that going to Trader Joe's is necessary for acquiring good chocolate (in an optimal way).

(6) A : I am looking for some good chocolate (to serve at the dinner).
B : You have to go to Trader Joe's.

On the standard analysis of modality due to Kratzer [13], the interpretation of the modal in (6) is determined by two contextual parameters, a modal base f and an ordering source g. We assume the analysis of goal-oriented modality in [4], which is based on a preferential ordering source (in line with [8]). In (6) it would consist of the goals of the addressee, including a goal to get good chocolate, inferred to motivate A's search.[1] The modal base is circumstantial, containing all the relevant facts, including, among others, the addressee's location and which stores do and do not carry good chocolate. (6) is then true in a world w iff in all worlds that agree with w on the relevant facts and which best satisfy A's goals at w, A goes to Trader Joe's.

[1] We will use the term 'goal' to refer to [4]'s notion of effective preference.

(2) has the same context-dependent truth conditions as (6). In order to incorporate the compositional contribution of the purpose clause, we assume that purpose clauses, like *if*-clauses, are modifiers of the modal, but operate on its ordering source parameter in the way specified in (7).

(7) A modal $\mathcal{M}_{f,g}$ modified by a purpose clause with content p is $\mathcal{M}^{p}_{f,g^+}$, where for any w, $g^+(w) = \{p \wedge r \mid r \in g(w)\}$

As long as g is non-empty and, for a world w, p is consistent with $g(w)$, which we can assume for the cases at hand, p will be true in all the f-g^+-best worlds relative to w.

The semantic content of (2) is that, relative to a world of evaluation, the f-g^+-best worlds, which are worlds in which you get good chocolate, are worlds in which you go to Trader Joe's. That in those worlds you get good chocolate *by* going to Trader Joe's and that, in the world of evaluation Trader Joe's carries good chocolate are pragmatic inferences. (See discussion in [14, 5] on how goal-oriented modal statements can convey factual information and the role of pragmatic reasoning in linking the prejacent to the stated goal as a way/means of achieving it.) The actual means implication is thus derived as an implication of goal-oriented necessity claims, generated on the basis of their semantic content and pragmatic reasoning.

2.2 Goal-oriented modality under *only*

We now turn to goal-oriented necessity clauses embedded under *only*. The null hypothesis is that 'To p, you only have to q' has the logical form in (8).

(8) $ONLY(\Box^{p}_{f,g}(q))$

Consider first a context in which graduate students applying for reimbursement for domestic travel need to obtain their advisor's signature, whereas international travel requires also signatures from the DGS and the chair. You ask about reimbursement for domestic travel, and receive the reply in (9).

(9) To apply for reimbursement, you only have to get the [advisor's]$_F$ signature.

(9) implies that your advisor's signature is required but no other signatures are. In this case, the alternatives are ranked by entailment and Horn's [9] account of the presupposition of *only* goes through without generating the prejacent problem. Assuming Rooth's [15] analysis, *only* is a propositional operator that brings into the semantic composition a contextually determined set of alternatives, based on the focus alternatives of the focused expression, including the prejacent proposition. *Only* presupposes that the prejacent alternative is true, and asserts that all the stronger alternatives are false.

Sufficiency modal constructions like (1) differ from (9) in that the alternatives are not ranked by strength. Accounting for them, therefore, requires moving away from entailment-based accounts of *only* to a more general scalar analysis (with entailment as a special case). Coppock & Beaver [6], building on Beaver & Clark [3], provide such an analysis. On their analysis, *only*'s semantics involves two operators, MIN and MAX, defined in (10). The presuppositional content of *only* is specified in terms of MIN and thus involves restricted existential quantification over alternatives. The assertive content of *only* is specified in terms of MAX. The inference to the prejacent arises from the combined effect of presupposition and assertion.

(10) a. $\text{MIN}(p, Alt) = \lambda w.\exists p' \in Alt(p)\,[p'(w) \wedge p \leq p']$
b. $\text{MAX}(p, Alt) = \lambda w.\forall p' \in Alt(p)[p'(w) \rightarrow p' \leq p]$

In (1), we can then take the focus alternatives to be stores that might, given the common ground between the interlocutors, carry chocolate, ordered, as assumed in the literature, by how easy they are to get to. Supposing the stores are your corner store, Trader Joe's, some other stores further away, and the distant chocolaterie *Coco*, the alternatives operated on by *only* would be as in (11), generated from the focus alternatives and inheriting their ordering. Due to the context dependence of the modal these alternatives are not propositions, but functions from contexts to propositions. What propositions they determine depends on how the global context resolves their pronoun-like parameters f and g^+.

(11) $\Box^{\text{choco}}_{f,g^+}$(you go to corner store) $<$ $\Box^{\text{choco}}_{f,g^+}$(you go to TJ) $<$ ... $<$ $\Box^{\text{choco}}_{f,g^+}$(you go to Coco)

Under what conditions could any of the necessities in (11) be true in a world in which multiple stores carry good chocolate? Our key assumption is that the ordering source g^+ in sufficiency modal constructions encodes a preference to minimize the cost of goal realization. Such an ordering source generally leads to a smaller domain of quantification than is involved in bare necessity statements like (2), and thus allows for a necessity to be true in a world in which multiple ways of achieving the goal exist. In any such world, only one of the alternative necessities will be true, namely the one in which the modal prejacent describes the least effortful way of getting chocolate, thus deriving the scalarity inference. Given that the contextual salience of a preference for cost minimization is a hallmark of the sufficiency modal construction, which singles out one way of achieving a goal as contrasting with higher ranked alternatives, this is a natural construal for the ordering source.

We model the preference for cost minimization by having the ordering source include propositions that the agent incurs no more than d-amount of cost, for different d's. This ensures that the ordering source contains propositions that stand in an entailment relation. On the resulting ordering on worlds, a world w_1 in which, for example, you get good chocolate incurring no more than d_1 amount of cost is better than a world w_2 in which you do so incurring no more than d_2 amount of cost, for $d_1 < d_2$, because w_1, in which you incur no more than d_1, and hence also no more than d_2 cost, verifies more propositions in the ordering source than w_2.

Under these assumptions, the presupposition in (12-a) is that the f-g^+-best worlds are ones in which you go to Trader Joe's or further and incur at least the amount d_{TJ} of cost associated with going to Trader Joe's. The assertion (12-b) is that the f-g^+-best worlds are ones in which you go no further than Trader Joe's and incur at most d_{TJ} cost. The combined presuppositional and assertive content is, therefore, that the f-g^+-best worlds are ones in which you go to Trader Joe's and incur d_{TJ}-cost, which is precisely the content our analysis assigns to *only*'s prejacent. As mentioned above, the conveyed content of (1) is then that in the best worlds, you get good chocolate incurring d_{TJ}-cost by going to Trader Joe's.

(12) a. P: $\lambda w.\exists p' \in Alt(\Box^{\text{choco}}_{f,g^+}(\text{you go to TJ}))\,[p'(w) \wedge \Box^{\text{choco}}_{f,g^+}(\text{you go to TJ}) \leq p']$
b. A: $\lambda w.\forall p' \in Alt(\Box^{\text{choco}}_{f,g^+}(\text{you go to TJ}))[p'(w) \rightarrow p' \leq \Box^{\text{choco}}_{f,g^+}(\text{you go to TJ})]$

(12) gives rise to the inference to the prejacent of *only*. Our main claim is that this is the correct result and straightforwardly explains the prejacent problem in terms of context equivocation (shifts in context between the premises and conclusions of an argument). Absent explicit modification, a natural context for a necessity clause like (2), unlike for a sufficiency modal construction, does not involve ranked alternative ways to achieve a goal, and hence determines an ordering source that does not encode cost minimization. This results in a larger domain of quantification and given the existence of multiple means, there is diversity across the worlds in the domain as to how the goal is achieved, falsifying the necessity clause.

This way of linking minimal sufficiency inferences to scalar alternatives and the ordering source has two further advantages. We correctly predict that necessity clauses can, in the right context, give rise to minimal sufficiency inferences in the absence of *only*. The implication in (13) is that flying across Europe is the least I have to do to see a Picasso.[2]

(13) To see a Picasso, you only have to drive downtown. I have to fly across Europe.

Cases like (13) provide empirical support for attributing the weakening of necessity brought about by *only* to a narrowed domain of universal quantification for the modal, as on our analysis (and Franke's), and against analyses that associate a possibility interpretation with sufficiency modal constructions, such as those of vF&I or Alonso-Ovalle & Hirsch (AO&H). For the former, bare necessity clauses are never predicted to have minimal sufficiency readings. For the latter, since nothing negates the necessity of higher alternatives, nothing ensures that the modal prejacent is in fact a way of achieving the goal.

We also capture the fact, pointed out in [11] (and echoed by [7]), that a scalar element in a goal-oriented necessity clause can lead to minimal sufficiency inferences without *only*, as seen in the implication of (14).

(14) To avoid evening rush hour without driving in the dark, you have to leave at 4pm.
$\rightsquigarrow$ *4pm is the latest you can leave and still avoid rush hour and driving in the dark*

A context invoking alternative means, ranked by some measure of cost, can determine an ordering source that encodes cost minimization, giving rise to the minimal sufficiency interpretation.

3 Comparison with previous proposals

vF&I decompose *only* into NEGATION + EXCEPT, and allow for split scope, with the modal taking scope between them.[3] The logical form of (1) is (15). EXCEPT has an existential presupposition, triggered under the modal in (15). (1) presupposes that in every world in which you get good chocolate you go to some (potentially different) store. The assertive content is (16), which, combined with the presupposition, yields goal-oriented possibility, as in (17).

(15) $\neg\Box^{choco}(\text{EXCEPT}(\text{TJ})\ \lambda x.\ \text{you go to } x))$

(16) $\neg\Box^{choco}(\exists x \neq \text{TJ} : \text{you go to } x)$

(17) $\Diamond^{choco}(\text{you go to TJ})$

As discussed above and in [12], this analysis does not meet the scalarity desideratum. The decomposition and split scoping of *only* is, furthermore, problematic. AO&H show that outside of the construction, the presupposition associated with *only* (via EXCEPT) is too weak, and that *only* has no negative component in languages where it is monomorphemic. [4] Finally, Huitink points out in [10] that a possibility semantics fails to account for valid transitivity inferences.

For AO&H [1], *only* presupposes its prejacent, which optionally contains a covert scalar operator AT LEAST. AT LEAST presupposes that its own prejacent is the lowest alternative, on an effort-based ranking, and asserts that one of the alternatives is true. The presuppostional

[2] Although both sentences imply minimal sufficiency, only the first has an easiness implication. This implication is due to *only*, but neither our account nor any of the alternatives explains why.

[3] In French, Greek and other languages, the translational equivalent of the sufficiency modal construction contains all three components on the surface.

[4] There is also no structural evidence for a bimorphemic analysis of *only*.

and assertive content amounts to a possibiliy claim as in vF&I, and is thus subject to some of the same objections. Furthermore, (18) shows that overt *at least* is illicit in the construction, weakening the case for a covert operator with presumably the same semantics.

(18) #To get good chocolate, you only have to go at least to Trader Joe's.

Krasikova & Zhechev [12] recognize the essential scalarity of the construction, but their account uses a special-purpose compositional analysis of *only*, and moreover derives very weak semantics, relegating what appears to be core semantic content to implicatures. *Only* takes a measure function, the modal prejacent, and the modal (modified by the purpose clause) as separate arguments, so *only* has no prejacent. The modal prejacent is mapped to a degree of likelihood, taken to correlate with difficulty.[5] (1) presupposes that in the closest worlds to the world of evaluation[6] in which the stated goal is achieved, you do something, and asserts that what you do is at least as likely (easy) as going to Trader Joe's. The combined content is that in the closest goal-worlds you do something at least as likely (easy) as going to Trader Joe's. A quantity implicature then excludes more likely (easier) alternatives, generating the content that in the closest goal-worlds, you do something with the same degree of likelihood as going to Trader Joe's. That what you do *is* go to Trader Joe's is claimed to be a relevance implicature.

Franke's [8] analysis meets the three desiderata and we rely on many of its insights. His key assumptions are that goal-oriented necessity clauses have scalar and non-scalar interpretations, and that on the scalar one they involve an ordering source that encodes cost minimization. For instance, (2) is claimed to allow an interpretation like (19-a) or (19-b).

(19) a. To get good chocolate, you have to go to Trader Joe's, if not even to Choco.
b. To get good chocolate, you have to go to Trader Joe's or further away.

He assumes further that *only* selects for the scalar interpretation,[7] and presupposes its prejacent. The presupposition of (1) then amounts to a disjunction of necessities only one of which is true in any given world, avoiding the prejacent problem. Goal-oriented clauses, however, do not generally have scalar, *at least*-type readings without overt modification. The contrast in felicity between (20-a) and (20-b) shows that necessities like (2) cannot actually be interpreted like (19-a) or (19-b). Moreover, as seen in (21), *only* is incompatible with overtly marked weak prejacents, undermining the hypothesis that *only* selects for scalar prejacents.

(20) a. To see a Picasso, you have to go at least to Chicago, but I'm not sure how far.
b. #To see a Picasso, you have to go to Chicago, but I'm not sure how far.

(21) a. #To get good chocolate, you only have to go to Trader Joe's if not even to Choco.
b. #To get good chocolate, you only have to go to Trader Joe's or further away.

Franke recognizes that goal-oriented necessity is context-dependent and that, in combination with *only*, it involves a particular contextual resolution of the ordering source, namely one that encodes cost minimization. He does not, however, connect this context dependence to the prejacent problem. What our analysis capitalizes on is that *only* requires a context that fixes the ordering source parameter of the modal in such a way that leads to the inference to the prejacent of *only*. The prejacent problem then emerges as a mirage.

[5] The idea being that more difficult things are less likely to happen. We think that introducing likelihood into the semantics is mistaken, since the likelihood inferences expected on this analysis are unattested.

[6] A similarity-based analysis to the modality problematically equates (2) with *if you were to get good chocolate, you would go to Trader Joe's.*

[7] This achieves the same weakening of the prejacent of *only* as AO&H's AT LEAST operator without stipulating a silent operator.

References

[1] Luis Alonso-Ovalle and Aron Hirsch. Keep *only* strong. *Semantics and Linguistic Theory 28 (SALT 28)*, pages 251–270, 2018.

[2] Luis Alonso-Ovalle and Aron Hirsch. Keep *only* strong. 2021. To appear in *Semantics and Pragmatics*.

[3] David Beaver and Brady Clark. *Sense and Sensitivity: How Focus Determines Meaning*. Blackwell, Oxford, 2008.

[4] Cleo Condoravdi and Sven Lauer. Anankastic conditionals are just conditionals. *Semantics & Pragmatics*, 9(8):1–69, 2016.

[5] Cleo Condoravdi and Sven Lauer. Conditional imperatives and endorsement. In *Proceedings of NELS 47*, pages 185–204, 2017.

[6] Elizabeth Coppock and David Beaver. Principles of the exclusive muddle. *Journal of Semantics*, 31(3):371–432, 2014.

[7] Émile Enguehard. Minimal sufficiency readings of necessity modals. In *Proceedings of NELS 50*, 2020.

[8] Michael Franke. Teleological necessity and *only*. In Janneke Huitink and Sophia Katrenko, editors, *Proceedings of the 11th European School Summer School in Logic, Language and Information (ESSLLI) student session*, pages 14–25, 2006.

[9] Laurence Horn. A presuppositional analysis of *only* and *even*. In *Fifth Regional Meeting of the Chicago Linguistics Society (CLS 5)*, pages 98–107, Chicago, 1969. University of Chicago, Chicago Linguistics Society.

[10] Janneke Huitink. Analyzing anankastic conditionals and sufficiency modals. In Sylvia Blaho, Luis Vicente, and Mark de Vos, editors, *Proceedings of ConSOLE XIII*, pages 135–156, 2005.

[11] Sveta Krasikova. Sufficiency inference in anankastic conditionals. *Semantics and Linguistic Theory 20 (SALT 20)*, pages 91–108, 2010.

[12] Sveta Krasikova and Ventsislav Zhechev. You only need a scalar "only". In Christian Ebert and Cornelia Endriss, editors, *Proceedings of the Sinn und Bedeutung 10*, pages 199–209, 2006.

[13] Angelika Kratzer. The notional category of modality. In Hans-Jürgen Eikmeyer and Hannes Rieser, editors, *Words, Worlds, and Contexts. New Approaches in Word Semantics*, pages 38–74. de Gruyter, Berlin, 1981.

[14] Sven Lauer and Cleo Condoravdi. Preference-conditioned necessities: Detachment and practical reasoning. *Pacific Philosophical Quarterly*, 95(4):584–621, 2014.

[15] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1:75–116, 1992.

[16] Kai von Fintel and Sabine Iatridou. Anatomy of a modal construction. *Linguistic Inquiry*, 38(3):445–483, 2007.

A scalar analysis of polarity contrasts between exclusive modifiers*

Lucas Fagen

The University of Chicago, Chicago, Illinois, U.S.A.
lfagen@uchicago.edu

Abstract

The English exclusive *only* licenses weak NPIs like *any* and *ever*, but its lexical cousins *just*, *merely*, *exclusively*, and *solely* do not. This paper sketches an analysis in which *only*'s status as an NPI licensor follows from its capacity to order alternatives by entailment. NPIs are contradictory with *just* and *merely*, which order alternatives by rank, and *exclusively* and *solely*, which exclude unordered alternatives.

1 The puzzle

Exclusive modifiers, which in English include *only*, *just*, *merely*, *exclusively*, and *solely*, form a lexical class (Coppock and Beaver 2014), conveying that some proposition is true (the *prejacent*, (1-a)) and that alternatives to the prejacent are false (1-b). Famously, *only* licenses weak negative polarity items (NPIs) like *ever* and *any* in its nuclear scope. The other exclusives do not (2). This is surprising, since they all exclude alternatives.

(1) My cat Gertrude only/just/merely/exclusively/solely eats kibble.
   a. → Gertrude eats kibble
   b. → Gertrude does not eat alternatives to kibble

(2) a. Gertrude only ever eats kibble.
   b. # Gertrude just ever eats kibble.
   c. # Gertrude merely ever eats kibble.
   d. # Gertrude exclusively ever eats kibble.
   e. # Gertrude solely ever eats kibble.

To my knowledge, the contrast between *only* and *just* re NPIs was first noticed offhand by Wagner (2006). For experimental confirmation of the judgment, see Callahan-Kanik (2018). An analysis of (2) has so far eluded the literature.

2 Background

2.1 Exclusives

This paper assumes the scalar lexical entry schema of Beaver and Clark (2008) and Coppock and Beaver (2014) (CB), who argue that exclusives operate on a scale of alternatives $\geq_s$ consisting of propositional answers to the current Question under Discussion (CQ) (Roberts 1996/2012). Exclusives presuppose a lower-bounding MIN statement that some true answer to the CQ is at least as strong as the prejacent on the relevant scale, and assert an upper-bounding MAX statement that no true answer is stronger than the prejacent (3).

*I thank Chris Kennedy, Anastasia Giannakidou, Itamar Francez, and three anonymous AC reviewers for feedback and discussion.

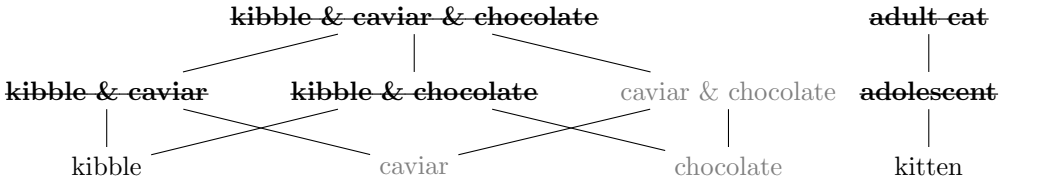
(3) a. $\text{MIN}_s(p) = \lambda w.\exists q_{\in \text{CQ}_s}[q(w) \wedge q \geq_s p]$
b. $\text{MAX}_s(p) = \lambda w.\forall q_{\in \text{CQ}_s}[q(w) \rightarrow p \geq_s q]$

Variation between exclusives is captured via variation in scale structure. Exclusives can have "complement exclusion" readings which exclude all other alternatives not entailed by the prejacent (4), and "rank-order" readings which only exclude alternatives ranked higher than the prejacent according to some contextual ordering (5).

(4) Gertrude only eats kibble. → Gertrude eats nothing **other** than kibble.

(5) Frederick is only a kitten. → Frederick is nothing **higher** than a kitten.

Complement exclusion readings rank the alternatives as a boolean lattice, closed under conjunction, so that recursively conjoined alternatives are included in addition to atomic ones: for example, *kibble&caviar* in addition to *kibble* and *caviar*. Logically stronger alternatives are ranked higher, so that the higher alternatives asymmetrically entail the lower ones. (4) is schematized in Figure 1a. Rank-order readings rank the alternatives as a qualitative ordering (Horn 2000) in which the higher alternatives can, but are not guaranteed to entail the lower ones. Since the alternatives do not necessarily stand in an entailment relation, they can be mutually exclusive. (5) is schematized in Figure 1b. The labels in Figure 1 are abbreviations for propositional alternatives; a ~~strikethrough~~ indicates exclusion.



(a) Entailment scale (b) Rank-order scale

Figure 1: Scales (adapted from CB ex. 27, 29.)

Complement exclusion sentences make exhaustive claims about all the alternatives on the scale. To see why, consider the scale in Figure 1a, which is ordered partially by entailment. MAX only excludes the boldfaced alternatives that entail the prejacent. But these collectively entail the other non-boldfaced alternatives (in this example, *caviar*, *chocolate*, and *caviar&chocolate*). So we can conclude from MIN and MAX that the non-boldfaced alternatives are false too. For example, if Gertrude eats kibble, and Gertrude does not eat kibble and caviar, then Gertrude does not eat caviar. When the higher alternatives do not entail the lower ones, rank-order sentences do not make exhaustive claims about everything on the scale. In particular, they say nothing about alternatives ranked lower than the prejacent.

Excluding along an entailment scale is truth-conditionally equivalent to excluding on non-identity, as the classic analysis of *only* (Horn 1969, Rooth 1992) had it. The motivation for introducing entailment scales is to tie complement exclusion and rank-order readings to a unified entry for *only*, which excludes along a scale whose structure is underspecified. In CB's system, the scale is provided by the context, not the exclusive, but different exclusives are compatible with different scales. According to CB, *only* "prefers" entailment scales and *just* "prefers" rank-order scales, while *merely* specifically presupposes an "evaluative" rank-order scale whose alternatives are ranked according to what the speaker considers good or bad. By contrast, *exclusively* and *solely* only have complement exclusion readings (6), which CB take to indicate that they are restricted to entailment scales. Their motivation for using entailment scales with

exclusively and *solely* is to achieve a unified scalar exclusive typology; non-scalar entries would capture the data just as well.

(6) Frederick is only/just/merely/#exclusively/#solely a kitten.

An exclusive's choice of scale indirectly correlates with whether the prejacent projects under negation (Klinedinst 2005). In negated complement exclusion sentences, it must; in negated rank-order sentences, it does not necessarily. (7) presupposes that an alternative at least as strong as the prejacent is true, and asserts that it is not the case that no stronger alternatives are true. It follows that at least one stronger alternative is true, and since these entail the prejacent, it survives negation. Gertrude still eats kibble in this example. (8) presupposes that an alternative ranked at least as high as the prejacent is true, and asserts that it is not the case that no higher alternatives are true. It follows that at least one higher alternative is true, but these do not entail the prejacent; in (8) they contradict the prejacent. So Frederick is no longer a kitten in this example.

(7) Gertrude doesn't only eat kibble (she eats chocolate). → Gertrude eats kibble.

(8) Frederick isn't only a kitten (he's an adult cat). ↛ Frederick is a kitten.

However, scale structure is not perfectly correlated with entailment or projection, because rank-order scales can still include atomic alternatives that entail other alternatives. It is ambiguous whether a scale like ⟨*employee, senior employee, CEO*⟩ has been ranked by entailment or seniority. Or consider Horn scales like ⟨*some, many, most, all*⟩ (Horn 1972). By definition the higher alternatives entail the lower ones, but since the atomic alternatives are not logically independent, it does not make sense to say that a Horn scale is organized as a boolean lattice.

Teasing apart entailment vs. rank is therefore not so straightforward. NPIs, which force complement exclusion readings in the scope of *only* (9), provide a good test. The absence of rank-order readings in (9-c) indicates that NPIs require scales ordered by entailment.

(9) *Context: card game* (example adapted from Klinedinst (2005) ex. 2)

a. I only/just/merely have a six. → Six is the **highest** card I have (rank-order)

b. Since the game started, I've only/just/merely had a six. → I have had no **higher** card than a six (rank-order)

c. Since the game started, I've only/#just/#merely ever had a six. → I have had no **other** card than a six (entailment)

This paper aims to derive (2) from scalar restrictions imposed by different exclusives.

2.2 NPIs

Let's assume with Chierchia (2013) that NPIs are existential quantifiers associated with maximally wide domains (10) that trigger obligatory exhaustification over domain alternatives (11). NPIs are grammatical when exhaustification does not contradict the prejacent (12). In (12-a), the output of *exh* is the proposition that there is no wider domain of time in which Gertrude doesn't eat kibble, which is entailed by the prejacent. In (12-b), the output of *exh* is the proposition that there is no narrower domain of time in which Gertrude eats kibble, which contradicts the prejacent.

(10) $[\![\text{ever}]\!] = \lambda e.\exists i_{\subset ever'}[\tau(e) = i]$

(11) $[\![exh]\!] = \lambda p \lambda w.p(w) \wedge \forall q_{\in \text{ALT}(p)}[^a q(w) \rightarrow {}^a p \geq_s {}^a q]$

(12) a. *exh*[Gertrude doesn't ever eat kibble.]

b. # *exh*[Gertrude ever eats kibble.]

Two features of the implementation will be useful. *Exh* is defined on partial propositions: ${}^{a}p$ in the exhaustification conjunct is shorthand for the at-issue component of p. This means *exh* ignores presuppositions. *Exh* is also scalar, sensitive to the same orderings exclusives are. This allows exhaustification over alternatives ordered by relations other than entailment. So if $\geq_s$ is a rank-order scale, *exh* will exclude alternatives ranked higher than the prejacent. That *exh* in (11) is defined identically to CB's MAX is intended to capture the intuition that *exh* is an abstract version of *only*[1].

Related implementations of the same idea (e.g. Krifka 1995) would work just as well here as long they accommodate rank-order scales.

3 Proposal

In treating scalar restrictions on exclusives as "soft preferences", CB's system as it stands is too permissive to explain (2), predicting that the presence of NPIs in the scope of exclusives other than *only* would force complement exclusion readings (as in (9-c)). Instead, they are completely out. Two modifications to CB's typology are necessary. First, we need to restrict *just* and *merely* to rank-order scales. Second, we need to model *exclusively* and *solely* as excluding everything other than the prejacent, as the classic analysis of *only* had it, without ranking or conjoining the alternatives. Only *only* can order alternatives by entailment[2]. *Just* and *merely* order alternatives by rank. *Exclusively* and *solely* operate on unordered alternatives. Setting aside whatever other factors distinguish *just* from *merely* and *exclusively* from *solely*, this analysis is implemented by the lexical entries in (13).

(13) a. $[\![\text{only}]\!] = \lambda p \lambda w : \text{MIN}_s(p)(w).\text{MAX}_s(p)(w)$
b. $[\![\text{just/merely}]\!] = \lambda p \lambda w : \text{RANK}(\geq_s) \wedge \text{MIN}_s(p)(w).\text{MAX}_s(p)(w)$
c. $[\![\text{exclusively/solely}]\!] = \lambda p \lambda w : p(w).\forall q_{\in \text{ALT}(p)}[p \neq q \rightarrow \neg q(w)]$

The NPI facts are captured as follows. Sentences with exclusives and NPIs include at least two dimensions of alternatives: the exclusive's focus alternatives (F-ALT) and the NPI's domain alternatives (D-ALT). The global alternative set ALT is represented as the union of F-ALT, D-ALT, and all the D-ALTs for each F-ALT (14).

(14) $\text{ALT}(p) = \text{F-ALT}(p) \cup \text{D-ALT}(p) \cup \{\text{D-ALT}(q) | q \in \text{F-ALT}(p)\}$

Scalar exclusives (*only*, *just*, and *merely*) only exclude higher-ranked F-ALTs, but impose the same ordering on the entire ALT set: either all alternatives are ordered by rank, or all are ordered by entailment. There are no "mixed" orderings. Revised definitions for MIN and MAX restricting exclusion to F-ALT are given in (15).

(15) a. $\text{MIN}_s(p) = \lambda w.\exists q_{\in \text{F-ALT}}[q(w) \wedge q \geq_s p]$
b. $\text{MAX}_s(p) = \lambda w.\forall q_{\in \text{F-ALT}}[q(w) \rightarrow p \geq_s q]$

[1] Panizza and Chierchia (2019) propose an exclusive semantics that also involves exhaustification over exclusive sentences, but in which *exh* only sees logically stronger alternatives, treating rank-order alternatives as disjunctions ranked by entailment. For example, they would represent the scale in Figure 1b as $\langle \textit{kitten} \vee \textit{adolescent} \vee \textit{adult cat}, \textit{adolescent} \vee \textit{adult cat}, \textit{adult cat} \rangle$. This move to eliminate rank-order exclusives entirely would require an independent explanation of the NPI facts. This paper takes the opposite strategy, treating *exh* as scale-flexible.

[2] The key factor is not boolean structure per se, but whether the ordering is explicitly specified as entailment. These need to be distinguished primarily because *only* still licenses NPIs when excluding along a Horn scale (i).

(i) Only/#just/#merely [some]$_F$ cats ate any dinner, not all.

The input to *exh* in sentences with exclusives and NPIs is the MAX statement, where the alternative set for a MAX statement consists of equivalent MAX statements with narrower subdomains (16). *Exh* inherits the same ordering $\geq_s$ as MIN/MAX.

(16) $\text{ALT}(\text{MAX}_s(p)) = \{\text{MAX}_s(q) | q \in \text{D-ALT}(p)\}$

By modeling both dimensions of alternatives as a larger structured set, we can derive the NPI data. Restrictions on F-ALT structure affect D-ALT too, in ways that either avoid contradiction, or create it.

Let's start with *only* in (17). Since *only* ranks the alternatives by entailment, MAX reverses strength: if $q \rightarrow p$, then $\text{MAX}_s(p) \rightarrow \text{MAX}_s(q)$. Therefore *exh* is vacuous. Since there is no narrower domain in which the proposition that Gertrude eats at most kibble in that domain entails that Gertrude eats at most kibble in the widest domain, there are no stronger alternatives for *exh* to operate on, as desired. As a result, for each excluded F-ALT, all its respective D-ALTs are also declared false, but none of the prejacent's D-ALTs are. This is not a contradiction.

Ordered pairs like ⟨*kibble, ever*⟩ in (17) are abbreviations for alternative propositions; the notation is intended to highlight the two dimensions along which the alternatives vary. A ~~strikethrough~~ indicates exclusion.

(17) $[\![exh[\text{Gertrude only ever eats kibble}]]\!] = (\exists i_{\subset ever'}[\tau(eat(k)(g)) = i])(\lambda p \lambda w : \text{MIN}_s(p)(w).\ \text{MAX}_s(p)(w) \wedge \forall q_{\in \text{ALT}(\text{MAX}_s(p))}[q(w) \rightarrow (\text{MAX}_s(p) \rightarrow q)]])$

a. $\text{F-ALT}(p) = \{\langle kibble,\ ever\rangle, \langle \sout{kibble \& caviar,\ ever} \rangle, \langle \sout{kibble \& chocolate,\ ever} \rangle, ...\}$

b. $\text{D-ALT}(p) = \{\langle kibble,\ ever\rangle, \langle kibble,\ sometimes\rangle, \langle kibble,\ often\rangle ...\}$

c. $\text{ALT}(p) = \{\langle kibble,\ ever\rangle, \langle \sout{kibble \& caviar,\ ever} \rangle, \langle \sout{kibble \& chocolate,\ ever} \rangle, \langle kibble,\ sometimes\rangle, \langle \sout{kibble \& caviar,\ sometimes} \rangle, \langle \sout{kibble \& chocolate,\ sometimes} \rangle, \langle kibble,\ often\rangle, \langle \sout{kibble \& caviar,\ often} \rangle, \langle \sout{kibble \& chocolate,\ often} \rangle ...\}$

d. $\text{ALT}(\text{MAX}_s(p)) = \{\text{MAX}_s(\langle kibble,\ ever\rangle), \text{MAX}_s(\langle kibble,\ sometimes\rangle), \text{MAX}_s(\langle kibble,\ often\rangle) ...\}$

Just and *merely* break the entailment relation between NPIs and their domain alternatives, forcing D-ALT onto a rank-order scale. Unlike (17), the ranking in (18) is not tied to logical strength and is not reversed by MAX: if $q \geq_s p$, then $\text{MAX}_s(q) \geq_s \text{MAX}_s(p)$, since $\text{MAX}_s(q)$ allows higher true alternatives than $\text{MAX}_s(p)$. So although the negation of each F-ALT entails the negation of its respective D-ALTs, rank is unaffected, and the narrower D-ALTs are still ranked higher. *Exh* is sensitive to rank: (11) defined it as scalar precisely to accommodate exhaustification over nonentailment scales. As a result, *exh* is not vacuous, excluding the higher-ranked, narrower MAX alternatives. (18) says that Gertrude eats nothing higher than kibble in the widest domain, but there is no alternative domain in which Gertrude eats nothing higher than kibble (i.e., that Gertrude *does* eat things ranked higher than kibble in narrower domains). This is a contradiction.

(18) $\#[\![exh[\text{Gertrude just ever eats kibble}]]\!] = (\exists i_{\subset ever'}[\tau(eat(k)(g)) = i])(\lambda p \lambda w : \text{MIN}_s(p)(w).\ \text{MAX}_s(p)(w) \wedge \forall q_{\in \text{ALT}(\text{MAX}_s(p))}[q(w) \rightarrow \text{MAX}_s(p) \geq_s q])$

a. $\text{F-ALT}(p) = \{\langle kibble,\ ever\rangle, \langle \sout{caviar,\ ever} \rangle, \langle \sout{chocolate,\ ever} \rangle, ...\}$

b. $\text{D-ALT}(p) = \{\langle kibble,\ ever\rangle, \langle kibble,\ sometimes\rangle, \langle kibble,\ often\rangle ...\}$

c. $\text{ALT}(p) = \{\langle kibble,\ ever\rangle, \langle \sout{caviar,\ ever} \rangle, \langle \sout{chocolate,\ ever} \rangle, \langle kibble,\ sometimes\rangle, \langle caviar,\ sometimes\rangle, \langle chocolate,\ sometimes\rangle, \langle kibble,\ often\rangle, \langle caviar,\ often\rangle, \langle chocolate,\ often\rangle ...\}$

d. $\text{ALT}(\text{MAX}_s(p)) = \{\text{MAX}_s(\langle kibble,\ ever\rangle), \sout{\text{MAX}_s(\langle kibble,\ sometimes\rangle)}, \sout{\text{MAX}_s(\langle kibble,\ often\rangle)} ...\}$

Now consider *exclusively* and *solely*, which are not scalar. In the present system, the scalar/non-

scalar distinction is implemented as exclusion of F-ALT vs. ALT. In sentences with one dimension of alternatives, ALT and F-ALT are the same, so that complement exclusion *only* vs. *exclusively/solely* sentences have equivalent truth-conditions. In sentences with NPIs, *exclusively/solely* exclude the prejacent's D-ALTs too. (19) says that Gertrude eats kibble in the widest domain, but not in any other domain. This is a contradiction[3]. *Exh* is vacuous here, so I've omitted it from the representation.

(19) $\#[\![\text{Gertrude solely ever eats kibble}]\!] = (\exists i_{\subset ever'}[\tau(eat(k)(g)) = i])(\lambda p \lambda w : p(w).\forall q_{\in \text{ALT}(p)}[p \neq q \rightarrow \neg q(w)])$

a. F-ALT$(p) = \{\langle kibble,\ ever\rangle, \langle \sout{caviar,\ ever}\rangle, \langle \sout{chocolate,\ ever}\rangle, ...\}$

b. D-ALT$(p) = \{\langle kibble,\ ever\rangle, \langle \sout{kibble,\ sometimes}\rangle, \langle \sout{kibble,\ often}\rangle ...\}$

c. ALT$(p) = \{\langle kibble,\ ever\rangle, \langle \sout{caviar,\ ever}\rangle, \langle \sout{chocolate,\ ever}\rangle, \langle \sout{kibble,\ sometimes}\rangle, \langle \sout{caviar,\ sometimes}\rangle, \langle \sout{chocolate,\ sometimes}\rangle, \langle \sout{kibble,\ often}\rangle, \langle \sout{caviar,\ often}\rangle, \langle \sout{chocolate,\ often}\rangle ...\}$

This is a somewhat surprising result. Previous analyses of *only* as an NPI licensor (von Fintel 1999, Chierchia 2013, Xiang 2017) have assigned it entries equivalent to (13-c), which the present analysis reserves for *exclusively* and *solely*. Once we understand F-ALT and D-ALT as different dimensions of the same set, an entry like (13-c) overexcludes. Modeling some but not all exclusives as scalar allows us to distinguish between those that operate on ordered alternatives (*only*, *just*, and *merely*) and unordered alternatives (*exclusively* and *solely*).

A brief comparison to Strawson downward entailment (von Fintel 1999) may be useful. von Fintel defines SDE as downward entailment plus a definedness condition on presuppositions. Concretely, if Gertrude only eats kibble in the widest domain, then Gertrude only eats kibble in narrower domains, presupposing that Gertrude eats kibble in narrower domains. To evaluate whether an SDE inference goes through, two conditions must hold: i) the consequent needs to follow from the premise, as long as ii) the presuppositions of the premise and consequent are satisfied.

Just and *merely* sentences do not satisfy condition i): on a rank-order scale, *exh* excludes the consequent. *Exclusively* and *solely* sentences do not satisfy condition ii): *exclusively/solely* exclude the presuppositions of the consequent, which are by definition D-ALT alternatives. So the effects of SDE are reconstructed here.

4 Conclusion

This paper argued that polarity contrasts between English exclusives result from restrictions on how the alternatives are ordered. The analysis has implications for our typology of exclusives: *just* and *merely* are restricted to rank-order scales, while *exclusively* and *solely* are not scalar at all. Aside from the present case study, the proposal provides a framework for analyzing sentences with multiple alternative-sensitive expressions whose dimensions interact.

[3]It follows that *exclusively* and *solely* do not exclude "innocently" (Fox 2007), which may go some way toward explaining why they resist operating on Horn scales (i).

(i) a. I only/#exclusively/#solely ate [some]$_F$ of the cookies, not all.
b. The movie was only/#exclusively/#solely [good]$_F$, not excellent.
c. Gertrude only/#exclusively/#solely played with Frederick [or]$_F$ Bruno, not both.

In general, the analysis predicts *exclusively/solely* to require logically independent alternatives. Alternatives that entail other alternatives are bad because they lead to contradiction. Mutually exclusive alternatives (as in (6)) are bad because presupposing the prejacent already resolves the other open alternatives, leaving *exclusively/solely* nothing to exclude.

References

Beaver, David and Brady Clark (2008). *Sense and sensitivity: how focus determines meaning*. Chichester: Wiley-Blackwell.

Callahan-Kanik, Taletha (2018). "An experimental study of the differences between just and only: NPI licensing and modal scope". PhD thesis. CUNY.

Chierchia, Gennaro (2013). *Logic in grammar: polarity, free choice, and intervention*. Oxford: Oxford University Press.

Coppock, Elizabeth and David Beaver (2014). "Principles of the exclusive muddle". In: *Journal of Semantics* 31, pp. 371–432.

Fox, Danny (2007). "Free choice and the theory of scalar implicatures". In: *Presupposition and implicature in compositional semantics*. Springer.

Horn, Laurence (1969). "A presuppositional analysis of only and even". In: *Proceedings of CLS 5*.

— (1972). "On the semantic properties of logical operators in English". PhD thesis. UCLA.

— (2000). "Pick a theory (not just any theory)". In: *Negation and polarity: syntactic and semantic perspectives*. Ed. by Laurence Horn and Yasuhiko Kato. Oxford: Oxford University Press.

Klinedinst, Nathan (2005). "Scales and *only*". MA thesis. UCLA.

Krifka, Manfred (1995). "The semantics and pragmatics of polarity items". In: *Linguistic Analysis* 25, pp. 209–257.

Panizza, Daniele and Gennaro Chierchia (2019). "Just exhaustification: a 'two-stage' theory of exclusives". In: *Proceedings of Amsterdam Colloquium 22*.

Roberts, Craige (1996/2012). "Information structure in discourse: Towards an integrated formal theory of pragmatics". In: *Semantics and Pragmatics* 5.6, pp. 1–69. DOI: 10.3765/sp.5.6.

Rooth, Mats (1992). "A theory of focus interpretation". In: *Natural Language Semantics* 1, pp. 75–116.

von Fintel, Kai (1999). "NPI licensing, Strawson entailment, and context dependency". In: *Journal of Semantics* 16.2, pp. 97–148.

Wagner, Michael (2006). "Association by movement: evidence from NPI-licensing". In: *Natural Language Semantics* 14, pp. 297–324.

Xiang, Yimei (2017). "*Only*: an NPI-licenser and NPI-unlicenser". In: *Journal of Semantics* 16.2, pp. 97–148.

Predicate Connection and Informational Oddness

Diego Feinmann
Institut Jean Nicod, Département d'Etudes Cognitives, ENS, EHESS, CNRS, PSL Research University
diegofeinmann@gmail.com

Abstract

In this short paper, I present a set of minimal pairs that calls into question all current accounts of contradictoriness and redundancy. In addition, I present two generalisations—grounded in the novel notion of *predicate connection* (Feinmann 2022)—that provide a fresh perspective on these phenomena.*

1 The Data

The following minimal pairs are interesting:

(1)	a. The Eiffel Tower is in Paris, and it isn't in France.	[FALSE + Contradictory]
	b. The Eiffel Tower is in Paris, and Paris isn't in France.	[Just FALSE]
(2)	a. The Eiffel Tower is in Paris, and it is in France.	[TRUE + Redundant]
	b. The Eiffel Tower is in Paris, and Paris is in France.	[Just TRUE]

The pair in (1) is interesting because it exposes the fact that we do not understand what it is that makes a sentence *feel* contradictory. Indeed, (1)a and (1)b are truth-conditionally equivalent; thus, no global notion of contradiction (e.g. falsity in every possible world) can tell them apart. If local contexts are brought into the picture, (1)a and (1)b are also equivalent: in both cases, the second conjunct contradicts its local context (cf. Schlenker 2009: 34).[1]

The pair in (2) is interesting because it calls into question all existing accounts of redundancy. According to non-incremental accounts (e.g. Meyer 2013; Katzir and Singh 2014), both (2)a and (2)b are predicted to be redundant; however, only (2)a feels redundant. Incremental accounts (e.g. Stalnaker 1974, 1978; van der Sandt 1992; Schlenker 2009; a.o.) cannot distinguish between (2)a

* For helpful discussions, many thanks to Philippe Schlenker, Daniel Rothschild, Amir Anvari, Émile Enguehard, Janek Guerrini, Matt Mandelkern, Dan Hoek, Jacopo Romoli, Moysh Bar-Lev, and Paul Egré. Special thanks to Benjamin Spector, Emmanuel Chemla, Manuel Križ, and Keny Chatain for extensive comments on earlier drafts. All errors my own.

Grant acknowledgments: The research leading to these results received funding from ANR-19-CE28-0004 (ProbaSem) and the European Research Council, ERC Grant Agreement N◦ 788077 (Orisem, PI: Schlenker). Research was conducted at the Département d'Etudes Cognitives, École Normale Supérieure - PSL Research University. The Département d'Etudes Cognitives is supported by grants ANR-17-EURE-0017 FrontCog and ANR-10-IDEX-0001-02 PSL.

[1] Note that, in all the sentences in (1) and (2), the local context of the second conjunct entails that Paris is in France; and it does because the global context entails that Paris is in France (it's common ground that Paris is in France).

and (2)b either: in both sentences, the second conjunct is entailed by its local context; as a result, both sentences are predicted to be redundant.

2 Predicate Connection

Here I will propose an account of both (1) and (2) that has at its core a novel theoretical notion—namely, the notion of *predicate connection* (Feinmann 2022).[2] To a first approximation, it can be said that two predicates α and β are connected (in a clause) iff their meanings are entangled (in the said clause) in such a way that, even if α and β didn't mean what they in fact mean but something else, this entanglement would nonetheless persist. This (admittedly vague) formulation is made precise in (3).

(3) **Predicate Connection**[3]

Let α and β be two one-place predicates, μ a clause, C_μ the local context of μ, and $\mathcal{D}_e$ the set of all entities. *P* and *Q* are two one-place predicate variables, and *f* a variable over assignment functions from {*P*, *Q*} to $\mathcal{D}_{\langle s,\langle e,t\rangle\rangle}$. For any v_1, $v_2 \in \{0,1\}$,

α and β are (v_1,v_2)-connected in μ with respect to C_μ ('in μ/C_μ' for short)[4] iff...

(i) α and β are both constituents of μ,
(ii) α isn't dominated by β nor is β dominated by α,[5] and
(iii) μ′—a clause just like μ except that α has been replaced by *P* and β by *Q*—satisfies (a) and (b):[6]

a. $\exists f \exists w \in C_\mu\left([\![\mu']\!]^{w,f} = 1\right)$[7]
b. $\forall f \forall w \in C_\mu\left([\![\mu']\!]^{w,f} = 1 \rightarrow \exists x \in \mathcal{D}_e \text{ s.t. } [\![P]\!]^{w,f}(x) = v_1 \wedge [\![Q]\!]^{w,f}(x) = v_2\right)$.

Notation: I'll say that two predicates α and β are *connected* in μ/C_μ (without further qualification) if, for some $v_1, v_2 \in \{0,1\}$, they are (v_1,v_2)-connected in μ/C_μ—that is, if they are (1,1)-connected in μ/C_μ, (1,0)-connected in μ/C_μ, (0,1)-connected in μ/C_μ, or (0,0)-connected in μ/C_μ).

If turned into a recipe, (3) would go more or less like this: first, identify a clause that has two constituents of predicative type (call one of these constituents α and the other β); second, replace α by *P* and β by *Q* (by so doing, one makes sure that the calculation of connection facts doesn't depend on the meaning of α and β); third, check whether the impoverished clause (the clause in which α has been replaced by *P* and β by *Q*) satisfies (iii)a and (iii)b. (Stripping away the technical details, (iii)a says: 'You can find a way of replacing α and β so that the resulting clause is consistent with its local context'; (iii)b, in turn, says, 'For all replacements of α and β, the resulting clause entails [in its local context] that there is an entity of which the replacement of α is v_1 and the replacement of β is v_2'.)

---

[2] Beyond the terminological resemblance, this notion is unrelated to the notion of *connectedness* (e.g. Chemla, Buccola, and Dautriche 2019).
[3] I thank Emmanuel Chemla for suggesting this manner of presentation.
[4] The empirical arguments for why C_μ (the local context of μ), as opposed to C (the global context), ought to be used for calculating connection facts are given in Feinmann (2022: § 4.1). In the examples discussed in this paper, all the relevant connection facts occur in matrix clauses; thus, for these examples, it doesn't matter whether C_μ or C is used (i.e. $C_\mu = C$ when μ is the matrix clause).
[5] This condition is needed in order for (iii) to be applicable.
[6] I'm assuming the following (non-standard) intepretation rule: if γ is an element of {*P*, *Q*}, then, for any *w* and for any *f*, $[\![\gamma]\!]^{w,f} = f(\gamma)(w)$; if γ is not an element of {*P*, *Q*}, then, for any *w* and for any *f*, $[\![\gamma]\!]^{w,f} = [\![\gamma]\!]^{w}$. To avoid clutter, I am omitting *g*, the assignment function that deals with the 'real' (as opposed to the artificially introduced) variables. This omission is harmless.
[7] (iii)a's function is to prevent (iii)b from being satisfied trivially.

Take (1)a, for example. The predicates 'is in Paris' and 'isn't in France' are (1,1)-connected in (1)a/$C_{(1)a}$, viz. in the matrix clause of (1)a with respect to (1)a's local context (which is of course the global context): indeed, these predicates are constituents of (1)a and neither of them dominates the other—hence (i) and (ii) are satisfied; furthermore, 'The Eiffel Tower *P*, and it *Q*' satisfies (iii)a—i.e. there is an *f* and a *w* in $C_{(1)a}$ s.t. $[\![$The Eiffel Tower *P*, and it *Q*$]\!]^{w,f} = 1$ (consider, for example, 'The Eiffel Tower *is in Paris* and it *isn't made of wood*'); and it also satisfies (iii)b—i.e. for any *f* and for any *w* in $C_{(1)a}$, if $[\![$The Eiffel Tower *P*, and it *Q*$]\!]^{w,f} = 1$, then there is an *x* in $\mathcal{D}_e$ s.t. $[\![P]\!]^{w,f}(x) = 1$ and $[\![Q]\!]^{w,f}(x) = 1$ (namely, the object denoted by 'the Eiffel Tower'). Note that, according to (3), the predicates 'is in Paris' and 'in France' are also connected in (1)a/$C_{(1)a}$: the difference is that, while 'is in Paris' and 'isn't in France' are (1,1)-connected in (1)a/$C_{(1)a}$, 'is in Paris' and 'in France' are (1,0)-connected in (1)a/$C_{(1)a}$.

I am now in a position to put forward the generalisation in (4)a, which predicts the contrast in (1), and the generalisation in (4)b, which predicts the contrast in (2).

(4) Let S be a sentence, μ a clause of S, C the global context (the context in which S is uttered), and C_μ the local context of μ.

(a) **Contradictoriness**

i. S is perceived as contradictory in C iff, for some $v_1, v_2 \in \{0,1\}$, there are two predicates that are (v_1,v_2)-connected in μ/C_μ and these two predicates are (v_1,v_2)-incompatible in C_μ.

ii. Two predicates α and β are (v_1,v_2)-incompatible in C_μ iff, for every $w \in C_\mu$, $\{x \in \mathcal{D}_e : [\![\alpha]\!]^w(x) = v_1\} \cap \{x \in \mathcal{D}_e : [\![\beta]\!]^w(x) = v_2\} = \emptyset$.

(b) **Redundancy**

i. S is perceived as redundant in C iff, for some $v_1, v_2 \in \{0,1\}$, there are two predicates that are (v_1,v_2)-connected in μ/C_μ and these two predicates are (v_1,v_2)-trivial in C_μ.

ii. Two predicates α and β are (v_1,v_2)-trivial in C_μ iff, for every $w \in C_\mu$, $\{x \in \mathcal{D}_e : [\![\alpha]\!]^w(x) = v_1\} \subseteq \{x \in \mathcal{D}_e : [\![\beta]\!]^w(x) = v_2\}$ or, for every $w \in C_\mu$, $\{x \in \mathcal{D}_e : [\![\beta]\!]^w(x) = v_2\} \subseteq \{x \in \mathcal{D}_e : [\![\alpha]\!]^w(x) = v_1\}$.

(1)a is expected to exhibit contradictoriness under (4)a: it has a clause that has two (1,1)-connected predicates that are (1,1)-incompatible, viz. 'is in Paris' and 'isn't in France'. (1)b, on the other hand, isn't expected to exhibit contradictoriness: this is because no clause of (1)b has two predicates that are connected; and, according to (4)a, having two connected predicates is a pre-requisite for contradictoriness. Likewise, (2)a is expected to exhibit redundancy under (4)b: it has a clause that has two (1,1)-connected predicates that are (1,1)-trivial, viz. 'is in Paris' and 'is in France'. (2)b, by contrast, is not expected to exhibit redundancy—for the same reason as (1)b is not expected to exhibit contradictoriness.

As explicitly stated, I take (4) to be generalisations, descriptions of the facts, and not theories, which should explain the facts and not just describe them. That said, I think it is possible to glean from (4), if not a theory, the scaffolding of one. Indeed, it is natural to conceptualise (4) along the following lines: connection facts impose constraints on what pairs of predicates that co-occur in a clause can mean; contradictoriness kicks in when these constraints are unsatisfiable; redundancy, in turn, kicks in when these constraints cannot not be satisfied.

3 Context and Informational Oddness

According to (4), context is expected to modulate our judgments of contradictoriness and redundancy; for example, a sentence may exhibit contradictoriness/redundancy in C_1, but not in C_2, because the connection facts that hold in C_1 do not hold in C_2. Consider, for example, (5) and (6): in the context in (5), the a-sentence exhibits contradictoriness (signalled with '^{c}') and the b-

sentence redundancy (signalled with '^{r}'); in the context in (6), by contrast, the sentences in question are oddness-free.

(5) [CONTEXT: it is common ground that Benjamin is a member of Linguae.]

a. ^{c} Every member of Linguae likes John, but Benjamin hates him.
b. ^{r} Every member of Linguae likes John, and Benjamin likes him too.

(6) [CONTEXT: it is common ground that Benjamin is a member of Parlare (not of Linguae).]

a. Every member of Linguae likes John, but Benjamin hates him.
b. Every member of Linguae likes John, and Benjamin likes him too.

The thing to note is that in (5), but not in (6), the global context entails that Benjamin is a member of Linguae: this difference alters the connection facts and, ultimately, the contradictoriness/ redundancy facts. Let's see this with (5)a/(6)a ('Every member of Linguae likes John, but Benjamin hates him'):

(5)a || $C_{(5)a}$ entails that Benjamin is a member of Linguae.

For any f and for any world $w \in C_{(5)a}$, if $[\![$Every member of Linguae P, but Benjamin $Q]\!]^{w,f} = 1$, then $[\![P]\!]^{w,f}(\text{Benjamin}'_w) = 1$ and $[\![Q]\!]^{w,f}(\text{Benjamin}'_w) = 1$.

'likes John' and 'hates (John)' are thus (1,1)-connected in (5)a/$C_{(5)a}$ (via Benjamin); contradictoriness is expected because 'likes John' and 'hates (John)' are (1,1)-incompatible in $C_{(5)a}$.

(6)a || $C_{(6)a}$ entails that Benjamin *is not* a member of Linguae.

It's not the case that, for any f and for any world $w \in C_{(6)a}$, if $[\![$Every member of Linguae P, but Benjamin $Q]\!]^{w,f} = 1$, then one of the following obtains:

- $[\![P]\!]^{w,f}(\text{Benjamin}'_w) = 1$ and $[\![Q]\!]^{w,f}(\text{Benjamin}'_w) = 1$
- $[\![P]\!]^{w,f}(\text{Benjamin}'_w) = 0$ and $[\![Q]\!]^{w,f}(\text{Benjamin}'_w) = 0$
- $[\![P]\!]^{w,f}(\text{Benjamin}'_w) = 1$ and $[\![Q]\!]^{w,f}(\text{Benjamin}'_w) = 0$
- $[\![P]\!]^{w,f}(\text{Benjamin}'_w) = 0$ and $[\![Q]\!]^{w,f}(\text{Benjamin}'_w) = 1$

'likes John' and 'hates (John)', therefore, are not connected in (6)a/$C_{(6)a}$ via Benjamin (nor are they connected in any other way); hence, these predicates aren't expected to induce contradictoriness.

According to (4), it's also possible that a sentence may exhibit contradictoriness/redundancy in C_1, but not in C_2, despite the connection facts in C_2 being the same as in C_1: this can happen when the incompatibility/triviality facts that hold in C_1 do not hold in C_2. Consider, for example, (7) and (8).

(7) [CONTEXT: it's common ground that the city of Tajiff is in Cuba.]

a. ^{c} John visited Tajiff, but didn't visit Cuba.
b. ^{r} John visited Tajiff, and he visited Cuba.

(8) [CONTEXT: it's common ground that the city of Tajiff is in Ecuador.]

a. John visited Tajiff, but didn't visit Cuba.
b. John visited Tajiff, and he visited Cuba.

In (7), but not in (8), the global context entails that the city of Tajiff is in Cuba. This difference has no impact on the connection facts: 'visited Tajiff' and 'didn't visit Cuba' are (1,1)-connected in both (7)a/$C_{(7)a}$ and (8)a/$C_{(8)a}$; and 'visited Tajiff' and 'visited Cuba' are (1,1)-connected in both (7)b/$C_{(7)b}$ and (8)b/$C_{(8)b}$. It does, however, have an impact on the incompatibility/triviality facts:

'visited Tajiff' and 'didn't visit Cuba' are (1,1)-incompatible in $C_{(7)a}$ but not in $C_{(8)a}$; 'visited Tajiff' and 'visited Cuba', in turn, are (1,1)-trivial in $C_{(7)b}$ but not in $C_{(8)b}$.

4 Mayr and Romoli's (2016) Datum

An attractive feature of (4)b is that, unlike previous theories, it is not threatened by (9), due to Mayr and Romoli (2016). Indeed, the sentence in (9) poses a serious challenge to both incremental and non-incremental accounts of redundancy: all these accounts predict (9) to be redundant, contrary to fact—see Mayr and Romoli (2016) for discussion.

(9) Either Mary isn't pregnant, or she is pregnant and it doesn't show.

(4)b fares better in this respect: it doesn't predict (9) to be redundant. Indeed, the predicates 'isn't pregnant' and '(is) pregnant', although they are (0,1)-trivial and (1,0)-trivial in $C_{(9)}$, are neither (0,1)-connected nor (1,0)-connected in (9)/$C_{(9)}$ (as a matter of fact, they aren't connected in any way); likewise, the predicates 'pregnant' and '(is) pregnant', although they are (1,1)-trivial and (0,0)-trivial in $C_{(9)}$, are neither (1,1)-connected nor (0,0)-connected in (9)/$C_{(9)}$.

Let's see this: To determine, for example, whether 'isn't pregnant' and '(is) pregnant' are (0,1)-connected in (9)/$C_{(9)}$, the following impoverished LF needs to be generated: 'Either Mary *P*, or she *Q* and it doesn't show'. If 'isn't pregnant' and '(is) pregnant' were to be (0,1)-connected in (9)/$C_{(9)}$, it would mean that, for any f and for any world $w \in C_{(9)}$, if $[\![$Either Mary *P*, or she *Q* and it doesn't show$]\!]^{w,f} = 1$, then there's an $x \in \mathcal{D}_e$ s.t. $[\![P]\!]^{w,f}(x) = 0$ and $[\![Q]\!]^{w,f}(x) = 1$. However, it's trivial to generate a counterexample to this: suppose that $w_1 \in C_{(9)}$ and that w_1 and f_1 are such that $\{x \in \mathcal{D}_e : [\![P]\!]^{w_1,f_1}(x) = 1\}$ = {Mary} and $\{x \in \mathcal{D}_e : [\![Q]\!]^{w_1,f_1}(x) = 1\} = \emptyset$. Then...

- $[\![$Either Mary *P*, or she *Q* and it doesn't show$]\!]^{w_1,f_1} = 1$;
- however, there's no $x \in \mathcal{D}_e$ such that $[\![P]\!]^{w_1,f_1}(x) = 0$ and $[\![Q]\!]^{w_1,f_1}(x) = 1$ (because there's no $x \in \mathcal{D}_e$ s.t. $[\![Q]\!]^{w_1,f_1}(x) = 1$).
- Hence, 'isn't pregnant' and '(is) pregnant' aren't (0,1)-connected in (9)/$C_{(9)}$.

Similar reasoning can be applied to show that 'isn't pregnant' and '(is) pregnant' aren't (1,0)-connected in (9)/$C_{(9)}$ either, and that 'pregnant' and '(is) pregnant' are neither (1,1)-connected nor (0,0)-connected in (9)/$C_{(9)}$.

It should be noted that, for the same reasons that (4)b predicts (9) not to be redundant, it also predicts (10) not to be redundant.

(10) Either Mary is pregnant, or she isn't (pregnant).

I take this to be a good prediction. This doesn't mean that (10) isn't problematic: this is a tautological sentence and, as a result, asserting it is expected to result in a violation of the maxim of Quantity (Grice 1975). However, as a sentence, (10) doesn't strike me as being redundant—that is, it doesn't strike me as exhibiting the same kind of oddness that 'Mary is pregnant, and she is pregnant' exhibits.

5 Hurford Disjunctions

(4)b doesn't predict Hurford disjunctions such as (11) to be redundant.

(11) #John was born in Paris or in France.

Non-incremental accounts might be seen as having an advantage here: these accounts rule out (11) essentially because one of its disjunct can be dropped without information loss. That said, these accounts fail to rule out examples such as (12), which are also deviant: this is because neither of (12)'s disjuncts can be dropped without information loss.

(12) #John was born in Russia or in Asia. (Singh 2008, attributed to Chomsky)

Thus, if one suspects that (11) and (12) are deviant for the same reason—a reasonable suspicion to have—then one is less likely to think that (11) is bad because one of its disjuncts is droppable.[8] As far as the account proposed here is concerned, the deviance of (11) and (12) is phenomenologically unrelated to the deviance of (2)a.

6 Closing Remarks

To conclude, I'd like to make a caveat: there at least two ways in which (4) is deficient. The first problem with (4) is that it fails to predict oddness in cases such as (13) below.[9]

(13) [c] John married Jane, then married Paula, but never married twice.

According to (4)a, in order for contradictoriness to kick in, two predicates need to be (v_1,v_2)-connected in μ/C_μ and, furthermore, (v_1,v_2)-incompatible in C_μ. In the example above, however, no such thing takes place. Take, for example, the matrix clause of (13): 'married Jane' and 'never married twice' are (1,1)-connected in (13)/$C_{(13)}$; however, these predicates aren't (1,1)-incompatible in $C_{(13)}$: it's possible for someone to have married Jane and, at the same time, not to have married twice. To deal with such cases, the definitions given need to be generalised to any number of predicates greater than 2 (see Feinmann 2022, § 4)—at the end of the day, what one wants to say is that (13) exhibits contradictoriness because 'married Jane', 'married Paula', and 'never married twice' are (1,1,1)-connected in (13)/$C_{(13)}$ and (1,1,1)-incompatible in $C_{(13)}$.

The other issue, brought to my attention by Benjamin Spector, is the following: it looks as if connecting one-place predicates isn't enough; to have a full description of the data, at the very least, one needs a system that, in addition to connecting one-place predicates, also connects generalised quantifiers (viewed, for this purpose, as higher-order predicates). Consider, for example, the following sentence:

(14) [c] None or all of the students speak French, and exactly half of them speak French.[10]

For (4)a to predict contradictoriness in (14), 'speak French' and 'speak French' would need to be either (1,0)-connected or (0,1)-connected in (14)/$C_{(14)}$ (i.e. 'speak French' and 'speak French' are both (1,0)-incompatible and (0,1)-incompatible in $C_{(14)}$). These predicates, however, are neither (1,0)-connected nor (0,1)-connected in (14)/$C_{(14)}$.[11] (4)a, therefore, makes a bad prediction here: it predicts (14) not to exhibit contradictoriness, contrary to fact.

Now, as Benjamin Spector pointed out to me, nothing stops one from having an account according to which the generalised quantifiers 'none or all of the students' and 'exactly half of (the students)' are (1,1)-connected in (14)/$C_{(14)}$;[12] on such an account, (14) would be expected to exhibit contradictoriness: 'none or all of the students' and 'exactly half of (the students)' are (1,1)-

[8] Thanks to Amir Anvari for discussion on this point.
[9] Thanks to Daniel Rothschild for making me reflect about these cases.
[10] Thanks to Benjamin Spector for this example.
[11] For a demonstration of this, see Feinmann (2022).
[12] Let P^2 and Q^2 be second-order predicate variables and k a variable over assignment functions from $\{P^2, Q^2\}$ to $\mathcal{D}_{\langle s,\langle\langle e,t\rangle,t\rangle\rangle}$; then, for any k and for any $w \in C_{(14)}$, if $[\![P^2 \text{ speak French and } Q^2 \text{ speak French}]\!]^{w,k} = 1$, then $[\![P^2]\!]^{w,k}([\![\text{speak French}]\!]^{w,k}) = 1$ and $[\![Q^2]\!]^{w,k}([\![\text{speak French}]\!]^{w,k}) = 1$.

incompatible in $C_{(14)}$ (i.e. there is no world w in $C_{(14)}$ and no one-place predicate denotation X such that $[\![\text{none or all of the students}]\!]^w(X) = 1$ and $[\![\text{exactly half of (the students)}]\!]^w(X) = 1$). To achieve this result, one would need to tweak (3) so that it also generates connection facts between generalised quantifiers, as well as upgrade the generalisations in (4).

The apparent fact that, in order to handle cases such as (14), connection needs to be established between expressions other than (first-order) one-place predicates raises the following question: how general does the definition of connection need to be? Should (3) be minimally modified to permit connection between generalised quantifiers (as suggested), or is, in fact, something more general needed (e.g. connection between any two expressions of type $\langle\langle ... \rangle, t\rangle$)? I leave this question open for future study.

References

Chemla, Emmanuel, Brian Buccola, and Isabelle Dautriche. 2019. "Connecting Content and Logical Words." *Journal of Semantics* 36 (3): 531–47.

Feinmann, Diego. 2022. "Contradictoriness." Manuscript, ENS Paris. https://ling.auf.net/lingbuzz/006629.

Grice, Herbert Paul. 1975. "Logic and Conversation." In *Syntax and Semantics: Speech Acts*, Peter Cole and Jerry L. Morgan, 3:41–58. New York: Academic Press.

Katzir, Roni, and Raj Singh. 2014. "Hurford Disjunctions: Embedded Exhaustification and Structural Economy." *Proceedings of Sinn Und Bedeutung* 18: 201–16.

Mayr, Clemens, and Jacopo Romoli. 2016. "A Puzzle for Theories of Redundancy: Exhaustification, Incrementality, and the Notion of Local Context." *Semantics and Pragmatics* 9 (0): 7–48. https://doi.org/10.3765/sp.9.7.

Meyer, Marie-Christine. 2013. "Ignorance and Grammar." Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/84420.

Schlenker, Philippe. 2009. "Local Contexts." *Semantics and Pragmatics* 2 (0): 3–78.

Singh, Raj. 2008. "On the Interpretation of Disjunction: Asymmetric, Incremental, and Eager for Inconsistency." *Linguistics and Philosophy* 31 (2): 245–60.

Stalnaker, Robert. 1974. "Pragmatic Presuppositions." In *Semantics and Philosophy*, edited by Munitz Milton K. and Peter Unger, 197–213. New York: New York University Press.

Stalnaker, Robert. 1978. "Assertion." In *Syntax and Semantics (New York Academic Press)*, 9:315–32. New York: New York Academic Press.

Van der Sandt, Rob A. 1992. "Presupposition Projection as Anaphora Resolution." *Journal of Semantics* 9 (4): 333–77.

Clausal embedding under TO in Japanese as speech acts*

Daniel Goodhue[1] and Junko Shimoyama[2]

[1] Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany
goodhue@leibniz-zas.de
[2] McGill University, Montreal, Quebec, Canada
junko.shimoyama@mcgill.ca

Abstract

We identify several puzzles, some novel, that are posed by clausal embedding under the Japanese complementizer TO. To resolve these puzzles, we present a unified syntax and semantics for TO and clausal embedding under it. Syntactically, we claim that TO-clauses are adjuncts. Interpretationally, we claim that TO embeds speech acts.

1 A syntax for *to* clauses

In (1), the TO-clause appears to be selected by the matrix attitude predicate *omou* 'think', and TO seems to be a standard declarative complementizer, similar to English *that*.

(1) Yoko$_k$-wa [**kanojo**$_k$-no jooshi-ga hannin da **to**] **omot**-te iru.
Yoko-TOP she-GEN boss-NOM culprit COP TO think-TE ASP.NPST
'Yoko thinks that her boss is the culprit.'

But, TO-clauses can also be 'unselected', as in (2), unlike English *that*-clauses (cf. [29, 9, 20]).

(2) Yoko$_k$-wa [**kanojo**$_k$-no uta-ga soto-ni kikoeru **to**] mado-o shimeta.
Yoko-TOP she-GEN song-NOM outside-DAT can.be.heard TO window-ACC closed
'Yoko closed the window, thinking that her singing can be heard from outside.'

Since TO-clauses do not need to be the complement of the matrix verb, they can appear in sentences in which another clause led by the complementizer KOTO serves as the complement:

(3) Yoko-wa [Sota-wa mada neteiru **to**] [kare-ga ookii oto-de ongaku-o kake-te
Yoko-TOP Sota-TOP still asleep TO he-NOM large volume-with music-ACC play-TE
iru **koto**]-o hiteishi-ta.
ASP.NPST KOTO-ACC deny-PST
'Yoko denied that Sota is playing loud music, saying/thinking that Sota is still asleep.'

Moreover, in some selected cases like (4), the propositional proform *soo* 'so' can appear in the complement of V, despite that, semantically-speaking, the attitude verb *shinjiru* 'believe' appears to get its propositional argument from the TO-clause. But if TO were replaced with KOTO as in (5), the result would be unacceptable with *soo*. The KOTO-clause appears to be the complement of V in way that the TO-clause is not.

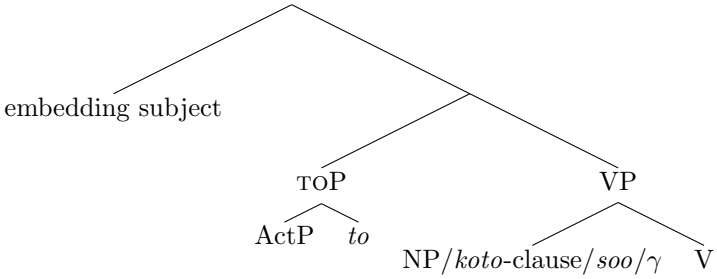
*For valuable comments, we thank Elizabeth Bogal-Allbritten, Fabrizio Cariani, Valentine Hacquard, Mako Hirotani, Keir Moulton, Paolo Santorio, Alexander Williams, the audiences at the Semantics workshop at Tokai, Ritsumeikan University, WAFL 14 at MIT, University of Ottawa, RelNomComp Workshop at the University of Toronto, GLOW in Asia 2019 in Seoul. The research reported here was supported in part by SSHRC (435-2015-0454 & 435-2022-0595 Moulton/Shimoyama) and by the ERC Advanced Grant 787929 "Speech Acts in Grammar and Discourse" (SPAGAD). All mistakes are our own.

(4) Yoko-no chiimu-ga katsu **to soo** shinji-te iru.
Yoko-GEN team-NOM win.NPST TO so believe-TE ASP.NPST
'(I) believe that Yoko's team will win.'

(5) Yoko-no chiimu-ga katsu **koto**-o (***soo**) shinji-te iru.
Yoko-GEN team-NOM win.NPST KOTO-ACC so believe-TE ASP.NPST
'(I) believe that Yoko's team will win.'

Especially in cases like (2)-(4), the TO-clause appears to be a VP adjunct. We adopt the VP adjunct analysis for all TO-clauses, and argue that this analysis helps to explain puzzling facts about cases where the TO-clause looks at first glance like it is the complement of an attitude verb. We present a schematic syntactic structure in (6) (cf. similar structures in [33, 9]).

(6) Schematic syntactic structure for sentences with TO-clauses



2 An interpretation for selected and unselected *to*-clauses

TO is sometimes described as a declarative complementizer (e.g. [18]). Others have treated it as a quotative marker (e.g. [26]). Given the unselected cases above, and other relevant properties below, we think the latter view is on the right track. However, while it is true that TO *can* embed direct quotations, it does not *have to*. First, notice that if (1) and (2) involved direct quotations, then the bolded pronouns coindexed with the matrix subjects would need to be first person, but they are not. Second, (7) demonstrates the ability to form matrix *wh*-questions from within the TO-clause, which would be impossible if this were a direct quotation.

(7) Yoko$_k$-wa [**kanojo**$_k$-no jooshi-ga **itsu** modoru to] ii-mashi-ta ka?
Yoko-TOP she-GEN boss-NOM when return.NPST TO say-POL-PST Q
'When did Yoko say that her boss would return?'

Thus we build on Saito's view, arguing that TO embeds *speech act phrases* (ActPs). This analysis will be crucial in explaining data from section 3. To implement this, we use an event semantics (building on [4, 9], a.o.). TO takes an ActP as input (we use the variable 'S' to range over speech acts), and acts as a speech report verb, as in (8).

(8) $[\![\text{to}]\!] = \lambda S.\lambda x.\lambda e.\lambda w.\ \exists e'[\text{utterance}(e', w)\ \&\ \text{agent}(e', x)\ \&\ \text{content}(e', w) = S\ \&\ e' \star e]$

We assume utterance events can be verbal or mental, and that their content is the content of the speech act (cf. [17]). (8) introduces an existentially bound event e' that is the event of the embedded speech act. The final conjunct pertains to how e' relates to the matrix event e. Note that '$\star$' cannot be mere temporal overlap: e and e' overlap temporally in (9), and yet it is infelicitous (cf. [10, 9]). We tentatively conclude that '$\star$' must enforce a stronger, causal relation between the two events.

(9) ??Yoko-wa [kyoo shokuba-de yat-ta koto-wa tadashikat-ta **to**] saba-o
Yoko-WA today work.place-at do-PST thing-WA right-PST TO mackerel-ACC
oobun-ni ireta.
oven-in put.in.PST
'Yoko put the mackerel in the oven, (while) thinking she did the right thing at work today.'

Given (6) and (8), the TOP and VP nodes will be of the same type, $\langle e\langle v\langle st\rangle\rangle\rangle$, thus combinable via predicate modification. Once combined with the matrix subject, and with the event existentially closed, the predicted interpretation for the unselected case in (2) will be as in (10). We leave ActPs unanalyzed here, as a full theory of speech acts is beyond the scope of this work. What matters is that speech acts are contentful events, attributable to an agent.

(10) $[\![(2)]\!] = \lambda w.\ \exists e[\text{closing}(e,w)\ \&\ \text{agent}(e,\textit{yoko})\ \&\ \text{patient}(e,\textit{the window})$
$\&\ \exists e'[\text{utterance}(e',w)\ \&\ \text{agent}(e',\textit{yoko})\ \&\ \text{content}(e',w) = [\![[_{ActP}$ kanojo-no uta-ga soto-ni kikoeru$]\!]\ \&\ e' \star e]]$

Turning now to our analysis of selected cases like (1), recall that the propositional argument of an attitude verb is sometimes delivered overtly by a KOTO-clause as in (3), or by a pronoun *soo* as in (4). When the propositional argument of the attitude comes from a TO-clause as in (1), we propose that the complement of V is a silent pronoun γ (as seen in (6)) that picks up its content from the speech act embedded in the TO-clause in the preceding part of the sentence (this analysis takes inspiration from the use of silent propositional anaphora in the focus literature, e.g. [25], as well as the polar particle literature, [13, 23]). Assuming a semantics for *omou* 'think' as in (11), the interpretation for (1) is as in (12).

(11) $[\![\text{omou}]\!] = \lambda p.\lambda x.\lambda e.\lambda w.\ \text{believe}(e,w)\ \&\ \text{exp}(e,x)\ \&\ \forall w' \in \text{content}(e,w)[p(w') = 1]$

(12) $[\![(1)]\!] = \lambda w.\ \exists e[\text{believe}(e,w)\ \&\ \text{exp}(e,\textit{yoko})$
$\&\ \forall w' \in \text{content}(e,w)[\textit{her boss is the culprit in } w']$
$\&\ \exists e'[\text{utterance}(e',w)\ \&\ \text{agent}(e',\textit{yoko})\ \&\ \text{content}(e',w)$
$= [\![[_{ActP}$ kanojo-no jooshi-ga hannin da$]\!]\ \&\ e' \star e]]$

3 Two interpretational puzzles explained by our analysis

The utility of the silent pronoun analysis becomes more apparent when considering (13) and (14), which introduce another interesting empirical property: TO enables polar interrogatives marked by the Q-particle *ka* to be 'selected by' predicates that do not select for interrogatives, *antirogative* predicates, like *omou* 'think' and *kitaisuru* 'hope'. Despite the acceptability of (13) and (14), since *omou* is antirogative, it cannot combine with the polar question meanings of these TO-clauses. Intuitively, interrogatives embedded under TO convey a weaker meaning than declaratives embedded under TO, as in (1). We refer to this weakening effect as *hedging*, and we indicate it in the English translation of (13) via the epistemic modal 'might'.

(13) Yoko$_k$-wa [kanojo$_k$-no jooshi-ga hannin **ka to**] **omot**-te iru.
Yoko-TOP she-GEN boss-NOM culprit Q TO think-TE ASP.NPST
'Yoko thinks that her boss **might** be the culprit.'

(14), which contains a negation, also has a hedging effect relative to (1). At the same time, (14) is not as hedged as the embedded positive interrogative in (13). The examples form a scale

then, with the embedded declarative (1) the least hedged, the embedded positive interrogative (13) the most hedged, and the embedded negative polar interrogative (14) in the middle. This is why we translate the declarative as unmodalized, the positive interrogative with 'might', and the negated version with 'probably'.

(14) Yoko$_k$-wa [kanojo$_k$-no jooshi-ga hannin ja **nai ka to**] **omot**-te i-ru.
Yoko-top she-GEN boss-NOM culprit COP.WA NEG Q TO think-TE ASP-NPST
'Yoko thinks that that her boss is **probably** the culprit.'

To analyze these facts, we hypothesize that the pronoun γ is able to pick up the speaker bias associated with questions embedded under TO. For a positive polar question like in (13), the idea is that the matrix subject's mental utterance of the question is motivated by positive evidential bias of the sort discussed in the literature on biased questions, which takes this bias to be a pragmatic inference ([2, 24, 28, 30, 14]. We model the bias here as a modalized proposition, *that her boss might be the culprit.* We do not necessarily believe that the bias actually *is* a modalized proposition (though such an analysis could be given in principle); rather, bias is a pragmatic implicature, and a modalized proposition is a reasonably close approximation of that meaning that allows us to explore the compositional interpretation of the matrix clause. Following [32, 4, 1], when a modal is embedded under a representational attitude, the worlds made accessible by the attitude serve as the domain for the modal. *Might* then imposes existential quantification on that domain, leading to the following interpretation for (13):

(15) $[\![(13)]\!] = \lambda w.\ \exists e[\text{believe}(e, w)\ \&\ \exp(e, yoko)$
$\&\ \exists w' \in \text{content}(e, w)[\textit{her boss is the culprit in } w']$
$\&\ \exists e'[\text{utterance}(e', w)\ \&\ \text{agent}(e', yoko)\ \&\ \text{content}(e', w)$
$= [\![[_{ActP}\ \text{kanojo-no jooshi-ga hannin ka?}]]\!]\ \ \&\ e' \star e]]$

The only difference (besides the content of the embedded speech act) is in the force of quantification over of the doxastically accessible worlds made available by the matrix attitude *omou* (existential in (15), universal in (12)). This difference in strength captures the intuitive difference discussed above, thus providing the desired interpretation for (13).

As for (14), the fact that it falls in the middle of the hedging scale between (1) and (13) as discussed above is unsurprising on the view that what is embedded in (14) is a positively biased negative polar question. High negation questions have been show to convey a relatively strong positive speaker bias in various languages ([24, 3]). Positive speaker bias in negative polar questions has been studied in Japanese ([28, 5, 8, 7, 19, 27]). If the embedded TO-clause in (14) contains a positively biased negative polar question, and if γ is picking up the biases of the questions embedded under TO, then we expect (14) to convey a hedge relative to (1), but less of a hedge than (13). We represent the stronger bias of negative polar questions in the metalanguage by a stronger, but non-maximal, modal like *probably* or *good possibility*, as analyzed in the graded modality of [11, 12] (the graded distinction could just as well be cast in terms of a probability semantics for modals, as in [15]). The interpretation of (14) then requires that there is a world w' among the doxastically accessible worlds such that the propositional prejacent of the question holds in all of the accessible worlds w'' more optimal than w'.[1]

[1] Following [12], the context c provides the function g, which takes the world of evaluation w as input, and produces the ordering source necessary to induce an ordering on the accessible worlds (being lower on the ordering means being more optimal):

(i) $\forall w, w' \in \text{content}(e, w)[w \leq_{g(w)} w' \Leftrightarrow \{p \in g(w) \mid w \in p\} \supseteq \{p \in g(w) \mid w' \in p\}]$

(16) $[\![(14)]\!] = \lambda w.\ \exists e[\text{believe}(e, w)\ \&\ \exp(e, yoko)$
$\&\ \exists w' \in \text{content}(e, w)$
$[\forall w'' \in \text{content}(e, w)[w'' \leq_{g(w)} w' \rightarrow \textit{her boss is the culprit in } w'']]$
$\&\ \exists e'[\text{utterance}(e', w)\ \&\ \text{agent}(e', yoko)$
$\&\ \text{content}(e', w) = \text{content}([\![[_{SAP}\ \text{kanojo}_k\text{-no jooshi-ga hannin ja nai}_{tc}\ \text{ka?}]]\!])\ \&\ e' \star e]]$

The result is as desired: the interpretation for (14) in (16) is stronger than the interpretation for (13) in (15) since, if every accessible w'' that is at least as optimal as a specific w' is one in which her boss is the culprit, then there is an accessible world in which her boss is the culprit, namely w'. And the interpretation for (1) in (12) is stronger than (16) since it says every accessible world is one which her boss is the culprit, thus there is one such that every world at least as optimal as it is one in which her boss is the culprit. So we have derived the scale of strength among the arguments of *omou* in our examples: (from strongest to weakest) (1) > (14) > (13).

One final interesting empirical property is that a TO-clause combined with *kitaisuru* 'hope' produces a stronger attitude than a KOTO-clause combined with *kitaisuru*, or English *hope that*. Suppose Yoko has no idea if her boss is the culprit, but she hopes he is. In such a context, the English sentence 'Yoko hopes that her boss is the culprit' is felicitous and true, but the Japanese sentence in (17) is not because it implies that Yoko takes the prejacent to be a much stronger likelihood than having no idea either way, as indicated by the English translation.

(17) Yoko_k-wa [kanojo_k-no jooshi-ga hannin da **to**] **kitaishi**-te ir-u.
Yoko-top she-GEN boss-NOM culprit COP.NPST TO hope-TE ASP-NPST
'Yoko hopes **and also strongly suspects** that her boss is the culprit.'

This generalization is further supported by the contrast in (18): The use of TO requires a high degree of confidence on the part of Yoko that the rain will stop, which in turn makes the continuation 'she's now almost given up' inappropriate. The acceptability of KOTO in (18) shows that these facts cannot be captured by proposing a stronger semantics for Japanese *kitaisuru* than English *hope*.

(18) Yoko-wa [ame-ga yam-u {**koto**-o/**??to**}] kitaishi-te-wa i-ru-kedo,
Yoko-WA rain-NOM stop-NPST KOTO-ACC/TO hope-TE-WA ASP-NPST-though
moo hotondo akirame-te i-ru.
already almost give.up-TE ASP-NPST
'Though Yoko hopes that the rain will stop (??and she strongly suspects it will), she's now almost given up.'

To explain these facts, we need a semantics for *kitaisuru*. The following is a simplified semantics for emotive doxastics based on [1] (see also [6, 31, 22, 21]).

(19) $[\![\text{kitaisuru}]\!] = \lambda p.\lambda x.\lambda e.\lambda w.\ \text{hope}(e, w)\ \&\ \exp(e, x)$
$\&\ \neg\forall w' \in \text{content}(e, w)[p(w') = 1]$ uncertainty requirement
$\&\ \exists w' \in \text{content}(e, w)[p(w') = 1]$ doxastic requirement
$\&\ p >_{DES_{x,w}} \neg p$ preference requirement

The content of a hoping event is a set of doxastically accessible worlds. Thus the uncertainty requirement in (19) is that the prejacent p doesn't hold throughout the doxastically accessible worlds (p is not maximally believed). The doxastic requirement in (19) is that the prejacent p

is doxastically possible (not believed to be false). The preference requirement in (19) is that p is preferred to $\neg p$ by x in w.[2] The predicted interpretation for (17) is as follows:

(20) $[\![(17)]\!] = \lambda w.\ \exists e[\text{hope}(e, w)\ \&\ \text{exp}(e, yoko)$
$\&\ \neg\forall w' \in \text{content}(e, w)[\textit{her boss is the culprit in } w']$ uncertainty requirement
$\&\ \exists w' \in \text{content}(e, w)[\textit{her boss is the culprit in } w']$ doxastic requirement
$\&\ \textit{her boss is the culprit} >_{DES_{x,w}} \neg\ \textit{her boss is the culprit}$ preference requirement
$\&\ \exists e'[\text{utterance}(e', w)\ \&\ \text{agent}(e', yoko)$
$\&\ \text{content}(e', w) = \text{content}([\![[_{SAP}\ \text{kanojo}_k\text{-no jooshi-ga hannin da}]]\!])\ \&\ e' \star e]]$

The asymmetry between the Japanese (17) and the English translation of it is that the prejacent merely needs to be doxastically possible in English, but that Yoko must believe it to be a stronger likelihood than that in Japanese. The semantics in (19) doesn't capture this fact about Japanese, since it merely requires the prejacent to be doxastically possible. However, it would be hasty to revise the semantics for *kitaisuru* 'hope' for two reasons. First, if we change the complementizer in (17) from TO to KOTO, then the doxastic requirement for the prejacent weakens to something indistinguishable from English, a fact that was further confirmed by the minimal pair in (18). This suggests that the weak doxastic requirement in the semantics in (19) is on the right track for Japanese after all. Second, the contribution of the TO-clause in (20) can account for the asymmetry, since it attributes an assertion of "Her boss is the culprit" to Yoko. We assume a commitment-based view of assertion ([16]), which bears an indirect relationship with the agent's beliefs: in many cases, an agent's choice to commit to vindicate the truth of a proposition p will coincide with the agent's belief in p. But nothing requires this—the agent can commit to p, and so assert it, even if p is merely very likely according to their beliefs. We can model this as an entailment from an agent's commitment to their beliefs as follows:

(21) If A commits to p, then $\exists\mathcal{O}$ such that $\mathcal{O}$ is an optimal subset of A's doxastically accessible worlds & $\mathcal{O} \subseteq p$

This entailment swamps the doxastic requirement of (19), making the Japanese (17) doxastically stronger than its English translation, as desired. At the same time, whenever a TO-assertion is embedded under *kitaisuru* 'hope', the uncertainty requirement of (19) will force $\mathcal{O}$ to be a *proper* subset of A's doxastically accessible worlds. This explains why Yoko's attitude about p in (17) is doxastically stronger than the English translation, and stronger than if the complementizer were KOTO, while at the same time explaining why Yoko's doxastic attitude about p in (17) is not maximally strong.

4 Conclusion

We have provided a unified analysis of 'selected' and 'unselected' TO-clauses that produces empirically adequate, compositionally-derived interpretations for various empirical facts, including two novel observations: (i) TO-embedding of declaratives and biased polar interrogatives produces a scale of strength among attitude arguments; and (ii) that TO+*kitaisuru* is stronger than a vanilla semantics for 'hope' would predict.

[2] Definition of $>_{DES_{x,w}}$ [1, p. 20]:

(i) $\forall w, w', w'' \in \text{content}(e, w)[w' >_{DES_{x,w}} w'' \Leftrightarrow w'$ is more desirable to x in w than $w'']$

(ii) $\forall p, q \subseteq W[p >_{DES_{x,w}} q \Leftrightarrow \forall w'' \in q[\exists w' \in p[w' >_{DES_{x,w}} w'']]\ \&\ \exists w' \in p[\forall w'' \in q[w'' \not>_{DES_{x,w}} w']]]$

References

[1] Pranav Anand and Valentine Hacquard. Epistemics and attitudes. In *Semantics and Pragmatics*, volume 6, pages 1–59, 2013.

[2] Daniel Büring and Christine A. Gunlogson. Aren't positive and negative polar questions the same? Ms. UCLA/UCSC, 2000.

[3] Daniel Goodhue. Isn't there more than one way to bias a polar question? *Natural Language Semantics*, pages 1–32, in press.

[4] Valentine Hacquard. On the event relativity of modal auxiliaries. *Natural Language Semantics*, 18(1):79–114, Mar 2010.

[5] Yurie Hara, Shigeto Kawahara, and Yuli Feng. The prosody of enhanced bias in Mandarin and Japanese negative questions. *Lingua*, 150:92–116, 2014.

[6] Irene Heim. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics*, 9:183–221, 1992.

[7] Hitomi Hirayama. Discourse effects of biased questions in Japanese. In *Japanese/Korean Linguistics 25*. CSLI Publications, 2018.

[8] Satoshi Ito and David Y. Oshima. On two varieties of negative polar interrogatives in Japanese. In Michael Kenstowicz, Theodore Levin, and Ryo Masuda, editors, *Japanese/Korean Linguistics*, volume 23. CSLI Publications, 2016.

[9] Jooyoung Kim. *Embedding without a license?: typology of unselected embedded clauses*. PhD thesis, University of Delaware, 2018.

[10] Jooyoung Kim and Satoshi Tomioka. Two types of unselected embedded questions. In *Proceedings of the 31st West Coast Conference on Formal Linguistics*, pages 276–284, 2014.

[11] Angelika Kratzer. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 2:201–216, 1981.

[12] Angelika Kratzer. Modality. In Arnim von Stechow and Dieter Wunderlich., editors, *Semantics: An international handbook of contemporary research*, pages 639–650. de Gruyter, Berlin, 1991.

[13] Manfred Krifka. Response particles as propositional anaphors. In *The proceedings of SALT*, volume 23, pages 1–18, 2013.

[14] Manfred Krifka. Bias in commitment space semantics: Declarative questions, negated questions, and question tags. In *Proceedings of Semantics and Linguistic Theory*, volume 25, pages 328–345, 2015.

[15] Daniel Lassiter. *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford University Press, 2017.

[16] John MacFarlane. What is assertion? In Jessica Brown and Herman Cappelen, editors, *Assertion*. Oxford University Press, 2011.

[17] Emar Maier. The pragmatics of attraction. In Saka Paul and Michael Johnson, editors, *The Semantics and Pragmatics of Quotation*, volume 15 of *Perspectives in Pragmatics, Philosophy & Psychology*. Springer, 2017.

[18] Teruyuki Mizuno. Q-particles in embedded declaratives, mood, and clausal complementation. In Nicole Dreier, Chloe Kwon, Thomas Darnell, and John Starr, editors, *Proceedings of Semantics and Linguistic Theory (SALT)*, volume 31, pages 405–425, 2022.

[19] David Y. Oshima. The prosody of positively biased negative polar interrogatives in Japanese: Post-focal reduction or deaccenting? In *Proceedings of the Annual Meeting of the Chicago Linguistic Society (CLS)*, volume 53, pages 275–289, 2019.

[20] Deniz Özyıldız. Embedded clauses in Turkish: Different paths to composition. Handout for Talk at Workshop on Relativization, Nominalization and Complementation, University of Toronto, June 2019.

[21] Paul Portner. *Mood*. Oxford University Press, 2018.

[22] Paul Portner and Aynat Rubinstein. Mood and contextual commitment. In Anca Chereches, editor, *The Proceedings of SALT 22*, pages 461–487, Ithaca, NY, 2013. CLC Publications.

[23] Floris Roelofsen and Donka Farkas. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, 91(2):359–414, 2015.

[24] Maribel Romero and Chung-Hye Han. On negative *Yes/No* questions. *Linguistics and Philosophy*, 27(5):609–658, 2004.

[25] Mats Rooth. A theory of focus interpretation. *Natural Language Semantics*, 1(1):117–121, 1992.

[26] Mamoru Saito. Sentence types and the japanese right periphery. In Ǵunther Grewendorf and Thomas Ede Zimmermann, editors, *Discourse and grammar: from sentence types to lexical categories*, volume 112 of *Studies in Generative Grammar*, pages 147–176. Mouton de Gruyter, Berlin, 2012.

[27] Junko Shimoyama, Daniel Goodhue, and Masako Hirotani. Positively biased negative polar questions in Japanese and their embeddability. In Tatiana Bondarenko, Colin Davis, Justin Colley, and Dmitry Privoznov, editors, *Proceedings of the 14th Workshop on Altaic Formal Linguistics (WAFL 14)*, MIT Working Papers in Linguistics 90, pages 289–304, Cambridge, Mass., 2019. MITWPL.

[28] Yasutada Sudo. Biased polar questions in English and Japanese. In Daniel Gutzmann and Hans-Martin Gärtner, editors, *Beyond Expressives: Explorations in Use-Conditional Meaning*, Current Research in the Semantics/Pragmatics Interface (CRiSPI) 28, pages 275–295. Brill, Leiden, 2013.

[29] Satoshi Tomioka and Jooyoung Kim. A new embedding strategy: Purposeful questions in japanese and korean. In *Proceedings of WAFL 11*, 2016.

[30] Tue Trinh. How to ask the obvious: A presuppositional account of evidential bias in english yes/no questions. In Luka Crnič and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 2, pages 227–249. MITWPL 71, 2014.

[31] Elisabeth Villalta. Mood and gradability: an investigation of the subjunctive mood in Spanish. *Linguistics and Philosophy*, 31:467–522, 2008.

[32] Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

[33] Akitaka Yamada. Embedded moods in Japanese. In *Proceedings of the 158th Meeting of the Linguistic Society of Japan*, pages 267–273, 2019.

Bias in high negation questions as a quantity implicature in commitment space semantics*

Daniel Goodhue

Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin, Germany
goodhue@leibniz-zas.de

Abstract

I analyze high negation as scoping over a commitment operator in a Commitment Space Semantics, building on [9]. I propose that polar questions chart two possible future courses through the commitment space, each rooted in a different possible commitment by the addressee. This semantics enables speaker bias in high negation questions to be derived as a special Quantity implicature, in competition with their positive polar question alternatives.

1 Bias and high negation

English high negation questions (HNQs) like in (1) necessarily convey that the speaker is biased for the positive answer, i.e. the propositional prejacent of the question embedded under negation. Prior work identifies HNQs crosslinguistically (Spanish, Bulgarian, Korean, German, Hungarian, Greek, and Turkish Sign Language; [14, 6, 5, 3]).[1]

(1) S: Didn't Mo eat? $\rightsquigarrow$ S is biased for the proposition *that Mo ate*

Unlike HNQs, positive polar questions (PPQs) and low negation questions (LNQs) do not necessarily convey speaker bias. Based on examples in [4], the contexts in (2) and (3) establish that the speaker S has no expectations about the answer to the question. Despite this lack of bias, the PPQ is acceptable in (2) while the LNQ is acceptable in (3). Meanwhile, HNQs are unacceptable in (2) and (3) because they necessarily convey a bias that S lacks in the contexts.

(2) S wants to find her roommate Mo, but has no expectations about whether Mo is home or not. She walks in the front door, sees their mutual roommate A and asks:
   a. Is Mo home? (PPQ)
   b. #Isn't Mo home? (HNQ)

(3) S wants to find her roommate Mo, but has no expectations about whether Mo is home or not. She looks everywhere and can't find Mo. But S does find their mutual roommate A in the last room that she checks. S says to A:
   a. Is Mo not home? (LNQ)
   b. #Isn't Mo home? (HNQ)

This asymmetry between HNQs on the one hand and PPQs and LNQs on the other raises the question, why does high negation give rise to this speaker bias in polar questions? In particular, what role is the preposing of negation playing?

*This work is supported by the ERC Advanced Grant 787929 "Speech Acts in Grammar and Discourse" (SPAGAD). Thanks to Anton Benz, Brian Buccola, Naomi Francis, Manfred Krifka, Marvin Schmitt, Bernhard Schwarz, Tue Trinh, Kazuko Yatsushiro, and two anonymous reviewers for the Amsterdam Colloquium for comments and discussion. All mistakes are my own.

[1] '$\rightsquigarrow$' indicates an inference arising from an utterance, without implying any particular analysis of it.

To address the latter question, [4, §4.2] provides several empirical tests showing that HNQs lack a negation in their prejacent, of which I will demonstrate two. *Again* presupposes that the proposition denoted by its complement has happened before. Since presuppositions project out of questions, we can check to see if the presupposition contains negation.

(4) Did Mo not eat again?
presupposes: Mo did not eat before

(5) Didn't Mo eat again?
presupposes: Mo ate before

In the LNQ in (4), the negation can be part of the presupposition. But in the HNQ in (5), it cannot, which shows that the high negation is too high for *again* to scope above.

Until- and *for*-adverbials only combine with clauses that have durative rather than punctual aspect. Adding negation to a clause with punctual aspect creates durative aspect. Thus *until*- and *for*-adverbials are acceptable in (6). However, they are not in HNQs.

(6) a. Did Mo not discover Bo until 9?
b. Did the ball not hit the ground for two minutes?

(7) a. #Didn't Mo discover Bo until 9?
b. #Didn't the ball hit the ground for two minutes?

This data suggest that *until*- and *for*-adverbials cannot scope above high negation.

In light of the facts discussed so far, there are two disiderata driving my treatment of polar questions:

1. A location for high negation in HNQs that explains the absence of negation in their prejacent (see section 2).

2. A semantics for HNQs that makes them weaker than their PPQ alternatives (see section 3), which will be crucial for the derivation of speaker bias in section 4.

2 Commitment space semantics

The Commitment Space Semantics developed here builds on [9]. A *commitment state* (CSt) is a context set, a set of worlds compatible with all the mutually shared commitments in the common ground à la [16]. *Commitment spaces* (CS) are sets of CSts that model future developments of the current CSt. Since new commitments shrink CSts, smaller CSts are possible continuations, as in (8a). CSs are rooted in a set of their largest CSts, as in (8b).

(8) a. If CSts c, c' are in CS C, and $c' \subset c$, then c' is a possible continuation of c in C.
b. $\sqrt{C}$, the root of C, is defined as $\{c \in C \mid \neg\exists c' \in C\ [c \subset c']\}$ [9, p. 68]

In Commitment Space Semantics, logical forms include distinct syntactic layers representing aspects of speech act structure, as in the LF for a PPQ in (9) [9, p. 64ff.]. Tense Phrases (TP) denote propositions, functions from possible worlds to truth values that characterize sets of worlds. Commitment Phrases (ComP) house the commitment operator '$\vdash$', which when combined with the TP also denote propositions. Speech Act Phrases (ActP) house speech act operators like question '**?**', that when combined with the ComP produce functions from Commitment Spaces to Commitment Spaces, similar to other dynamic semantics.

(9) $[_{\text{ActP}}\ \text{Did}_k\ \textbf{?}\ [_{\text{ComP}}\ \vdash\ [_{\text{TP}}\ \text{Mo}\ t_k\ \text{eat}\ ]\ ]\ ]$ ('$[_{\text{TP}}$ Mo ate $]$' abbreviated below as ϕ)

Given these structural assumptions, high negation can be analyzed as scoping above '$\vdash$', as in (10). This makes good on disideratum 1, explaining why the tests in section 1 demonstrate a lack of negation in the prejacent of HNQs, on the assumption operators like *again* and *until*- and *for*-adverbials cannot scope into the speech act structure.

(10) $[_{\text{ActP}}\ \text{Did}_k\ \textbf{?}\ [_{\text{ComP}}\ \text{n't}\ \vdash\ [_{\text{TP}}\ \text{Mo}\ t_k\ \text{eat}\ (\phi)\ ]\ ]\ ]$

I model commitment as a kind of attitude predicate in (11). $\mathcal{R}_j$-accessible worlds are those that are compatible with the judge j's commitments in w. If j commits to a proposition p, it means j claims publicly that p is true, and that j will provide arguments supporting p if p were to be challenged [12, 1, 11, 8].

(11) $[\![\vdash]\!]^{i,\,d,\,g} = \lambda p.\lambda w.\ \forall w'[w\mathcal{R}_j w' \rightarrow p(w') = 1]$ (abbreviated as: $\lambda p.\lambda w.\vdash_{j,w} p$)

The interpretation function has three parameters, a set of the interlocutors i, a set of salient discourse referents d, and an assignment function g. j is a variable in the assignment function g whose value is fixed by whichever speech act operator is in ActP. An assertion operator in ActP shifts j to the speaker S in the g of its complement, while the question operator **?** shifts j to the addressee A in the g of its complement, in addition to having other effects that are elided for the moment in (12).

(12) $[\![\textbf{?}\ \alpha]\!]^{i,\,d,\,g} = \lambda C.\ \ldots\ [\![\alpha]\!]^{i,\,d,\,g[j\,\rightarrow\,A]}\ \ldots$

3 Interpretations for PPQs and HNQs

For Krifka, polar questions are usually assumed to restrict the CS to just those developments in which A makes the commmitment in ComP. E.g. (9) would produce $\sqrt{C}\ \cup \{c \in C \mid c \subseteq \lambda w.\vdash_{A,w} \phi\}$. This model leaves out CSts in which A makes a commitment other than the one in the ComP of the question, e.g. $\vdash_{A,w} \neg\phi$. I revise [9]'s denotation for the question operator '**?**' to output a multi-rooted CS that charts two opposing paths through the commitment space.

The idea for **?**'s semantics is intuitive enough: **?** takes an input ComP, and it projects two opposing future continuations of the conversation:

- Continuation 1: A makes the commitment $[\![[_{\text{ComP}}\ \alpha]]\!]$.
- Continuation 2: A makes a commitment to an alternative, $[\![[_{\text{ComP}}\ \alpha']]\!]$, that would be incompatible with $[\![[_{\text{ComP}}\ \alpha]]\!]$, i.e. $[\![[_{\text{ComP}}\ \alpha]]\!] \cap [\![[_{\text{ComP}}\ \alpha']]\!] = \varnothing$

But if the idea for **?** is intuitive, the implementation is not. The simplest way to find an alternative that would be incompatible with $[\![[_{\text{ComP}}\ \alpha]]\!]$ would be to negate it, i.e. $[\![[_{\text{ComP}}\ \alpha']]\!] = \neg[\![[_{\text{ComP}}\ \alpha]]\!]$. This would make **?** similar to a standard Q morpheme. However, given the LFs for PPQs in (9) and HNQs in (10), this is a nonstarter because the denotations of PPQs and HNQs would be identical, predicting no meaning differences between them and so failing to satisfy desideratum 2.

What is needed is for **?** to be able to 'see' the TP proposition, in addition to its complement ComP. Then, given the meaning $[\![[_{\text{ComP}}\ \alpha]]\!]$, **?** can produce the opposing meaning $[\![[_{\text{ComP}}\ \alpha']]\!]$ by either negating the TP or not, and then combining the result with a commitment operator $\vdash$:

(13) a. If $[\![[_{\text{ComP}}\ \alpha]]\!] = \lambda w.\vdash_{j,w} \phi$, then $[\![[_{\text{ComP}}\ \alpha']]\!] = \lambda w.\vdash_{j,w} \neg\phi$,
since $\lambda w.\vdash_{j,w} \phi\ \cap\ \lambda w.\vdash_{j,w} \neg\phi = \varnothing$

b. If $[\![[_{\text{ComP}}\ \alpha]]\!] = \lambda w.\neg \vdash_{j,w} \phi$, then $[\![[_{\text{ComP}}\ \alpha']]\!] = \lambda w.\vdash_{j,w} \phi$,
since $\lambda w.\neg \vdash_{j,w} \phi \cap \lambda w.\vdash_{j,w} \phi = \varnothing$

To give a semantics for '**?**' that achieves this, I assume that TPs produce propositional discourse referents DR_ϕ (cf. [7] & [13]) that are recorded in the parameter d, and that are accessed by the question operator **?**.[2] **?** has the syncategorematic semantics in (14). It is a function from an input CS C to an output CS restricted to those elements c in C that are subsets of either $[\![[_{\text{ComP}}\ \alpha]]\!]$ or $[\![[_{\text{ComP}}\ \alpha']]\!]$. The polarity operator variable 'y', based on an idea proposed by [15, p. 17]), is **?**'s means of either negating the TP DR_ϕ or not, depending on what $[\![[_{\text{ComP}}\ \alpha]]\!]$ is.

(14) $[\![\ \textbf{?}\ [_{\text{ComP}}\ \alpha]]\!]^{i,d,g} = \lambda C.\ \{c \in C \mid c \subseteq [\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]} \vee c \subseteq [\![\vdash]\!]^{i,d,g[j \to A]}(y(\text{DR}_\phi))\}$,
where y is a variable ranging over polarity operators—i.e. $y \in \{\lambda p.p, \neg\}$—set to whichever operator produces an alternative commitment incompatible with $[\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]}$, i.e.: $[\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]} \cap [\![\vdash]\!]^{i,d,g[j \to A]}(y(\text{DR}_\phi)) = \varnothing$.

Alternatively, we can think of **?** as a conditional function as in (15):

(15) $[\![\ \textbf{?}\ [_{\text{ComP}}\ \alpha]]\!]^{i,d,g} = \lambda C.\ \{c \in C \mid c \subseteq [\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]} \vee c \subseteq \ldots$
a. $[\![\vdash]\!]^{i,d,g[j \to A]}(\text{DR}_\phi)\}$, if $[\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]} \cap [\![\vdash]\!]^{i,d,g[j \to A]}(\text{DR}_\phi) = \varnothing$
b. $[\![\vdash]\!]^{i,d,g[j \to A]}(\neg\text{DR}_\phi)\}$, if $[\![[_{\text{ComP}}\ \alpha]]\!]^{i,d,g[j \to A]} \cap [\![\vdash]\!]^{i,d,g[j \to A]}(\neg\text{DR}_\phi) = \varnothing$

(14) or (15b) produces the CS in (16) for the PPQ in (9):[3]

(16) $[\![\text{Did Mo eat?}]\!]^{i,d,g} = \{c \in C \mid c \subseteq \lambda w.\vdash_{A,w} \phi \ \vee\ c \subseteq \lambda w.\vdash_{A,w} \neg\phi\}$

In producing (16), y is set to $\neg$ (put otherwise, (15b) is chosen) because $\lambda w.\vdash_{A,w} \phi \cap \lambda w.\vdash_{A,w} \neg\phi = \varnothing$, whereas if y had been set to $\lambda p.p$, the two propositions would have been identical.

To compare this with the HNQ in (10), consider the semantics for the HNQ's ComP in (17), and the resulting CS in (18):

(17) $[\![\ \text{n't}\ ]\!]^{i,d,g}\ ([\![\vdash\ [_{\text{TP}}\ \phi\ ]\ ]\!]^{i,d,g}) = \lambda p.\lambda w.\neg p(w)\ (\lambda w.\vdash_{j,w} \phi) = \lambda w.\neg \vdash_{j,w} \phi$

(18) $[\![\text{Didn't Mo eat?}]\!]^{i,d,g} = \{c \in C \mid c \subseteq \lambda w.\neg \vdash_{A,w} \phi \ \vee\ c \subseteq \lambda w.\vdash_{A,w} \phi\}$

To produce (18), y is set to $\lambda p.p$ (put otherwise, (15a) is chosen) because $\lambda w.\neg \vdash_{A,w} \phi \cap \lambda w.\vdash_{A,w} \phi = \varnothing$, whereas if y had been set to $\neg$, there would have been entailment relation between the two propositions.

Both (16) and (18) have two mutually incompatible root CSts that lead to incompatible continuations: Each has a root in which $\vdash_{A,w} \phi$; and while (16) has a $\vdash_{A,w} \neg\phi$ root, (18) has a $\neg \vdash_{A,w} \phi$ root. (14)/(15) achieves this by producing CSs that include the ComP commitment and its obvious alternative given the TP prejacent proposition.

[2] An alternative to assuming DR_ϕ in d would be to preserve DR_ϕ in the compositional semantics via an ordered pair representation of the TP such that $[\![[_{\text{TP}}\ \phi]]\!] = \langle\phi, \text{DR}_\phi\rangle$ (cf. [10, p. 6]). Composition rules would then be revised to operate on the first coordinate of the ordered pair, while the discourse referent is passed up the structure in the second coordinate.

[3] In (16) and the following, I overload the symbol 'ϕ': In the object language, 'ϕ' is an abbreviation for the TP constituent '$[_{\text{TP}}$ Mo ate $]$'. In the metalanguage, 'ϕ' is an abbreviation for the proposition denoted by that TP '$\lambda w.$ *Mo ate in* w'. I.e. $[\![\phi]\!] = \phi$. Moreover, DR_ϕ has the same meaning: $[\![\text{DR}_\phi]\!] = [\![\phi]\!]$.

4 Speaker bias as quantity implicature

Taking the familiar case of sentences with **all** vs. **some**, quantity implicature can be thought of as depending on the (defeasible) conditional in (19). If S says **some**, then the consequent doesn't hold, and by *modus tollens* S doesn't believe **all**, which can be strengthened into ¬**all**.

(19) If S believes **all**, then **all** is more useful to utter than **some** (because it is more informative).

Compare the preceding to the following derivation of speaker bias. I posit the conditional in (20):

(20) If S is ignorant of whether p or $\neg p$, then the PPQ with prejacent p (PPQ-p) is more useful than HNQ-p.

(20) depends on the assumption that if S is ignorant about some relevant p, then their goal is to gain information about it. (20) also depends on the claim that any PPQ-p is more informative than HNQ-p. The semantics in (16) and (18) deliver this: The CS in (16) is a proper subset of that in (18), and in particular, one of the root CSts of (16) is a proper subset of one of the root CSts of (18). This in turn depends on the simple fact that $\lambda w.\vdash_{A,w} \neg\phi$ entails $\lambda w.\neg \vdash_{A,w} \phi$. Relative informativity of questions is defined as follows:

(21) Q_1 is more informative than Q_2 iff the following two conditions are satisfied:
a. $\exists c \in \sqrt{Q_1}\ [\exists c' \in \sqrt{Q_2}\ [c \subset c']]$ b. $\forall c \in \sqrt{Q_1}\ [\neg\exists c' \in \sqrt{Q_2}\ [c' \subset c]]$

If S chooses to use HNQ-p instead of PPQ-p, it implies that PPQ-p is not more useful. Given (20) and *modus tollens*, this implies that S is not ignorant of whether p or $\neg p$. In other words, S is either biased for p or for $\neg p$. This bias implicature is not cancellable because whenever HNQ-p is relevant, the more informative PPQ-p alternative is also relevant, since both are about the same prejacent proposition p.

With this non-directional bias in hand, the last step is to explain why HNQ-p always conveys bias for p (as opposed to $\neg p$). This can be read off of the structure of the CS in (18). Suppose S were biased for $\neg\phi$. If A were to pursue the commitment in the first disjunct in (18), then S would know that S and A have a difference of opinion with respect to ϕ. But if A were to pursue the commitment in the second disjunct in (18), S would not yet know whether A agrees with S that $\neg\phi$. This is because A's lack of commitment to ϕ does not imply a commitment to $\neg\phi$; it is also compatible with a lack of commitment either way, for both ϕ and $\neg\phi$. Compare this with the case in which S is biased for ϕ. In that case, a choice by A to pursue the commitment in either disjunct in (18) would immediately reveal to S whether or not they are on the same page with respect to ϕ: the first disjunct means they agree about ϕ, the second that they do not. Thus if the goal of conversation is to grow the common ground via mutual commitment, then it only makes sense to use an HNQ when biased for the TP proposition that the HNQ is built from.

5 Comparisons to other approaches

One reviewer for the *AC* points out an asymmetry between the above account and [9]: [9] provides a monopolar analysis of polar questions so that they only project continuations in which A makes the commitment in the ComP, as mentioned above. One benefit of [9]'s approach is it produces distinct semantics for PPQs, LNQs, HNQs, and negative alternative questions

(NAQs) like *Did Mo eat or didn't she?*. As for the account I have developed, although I didn't provide a semantics for LNQs or NAQs, I agree with the reviewer that my semantics, without any further changes, predicts PPQs, LNQs, and NAQs to all have the same semantics (while the semantics for HNQs will of course be unique). Giving a complete account of LNQs and NAQs is beyond the scope of this paper, but I would aim to explain the unique properties of NAQs via the fact that both alternatives are overtly uttered, which makes each alternative equally salient (cf. discussion of salience in [7], and highlighting in [13]). As for the well known difference in evidential bias between PPQs and LNQs, I would similarly like to exploit the difference in the prejacents uttered, the overt negation in LNQs in particular, and I would aim to explain the negative evidential bias conveyed by LNQs via the markedness of negation (cf. discussion in [2, 18, 17, 13]).

Another reviewer wonders how HNQs relate to assertions on my account. While I don't make this claim here, in [4] I argue that the bias in HNQs may be as strong as full belief (at least in some cases). The reviewer asks why an S who believes p would use HNQ-p instead of asserting p. As I discuss in [4], HNQs are useful for determining agreement. S may have believed p up until a few moments before uttering the HNQ as the result of being confronted with evidence against p. Another possibility discussed in [4] is that HNQs are often reserved for use in contexts in which A has at least as much epistemic or social authority over whether or not p is true as S does. Such contexts make S more likely to use an HNQ, to simultaneously convey their bias for p while giving A control over whether p is determined to be true or not. Exploitation of this authority asymmetry may also be used to allow A to save face.

As for the strength of bias, my account does not require bias to be as strong as full belief. To achieve a weaker bias, (20) would need to be reconsidered. If the negation of ignorance of whether p or $\neg p$ is belief in either p or $\neg p$,[4] then to weaken the bias, the antecedent of (20) needs to be strengthened to something like 'complete ignorance', i.e. not only does the antecedent say that S is ignorant, but that S isn't even leaning one way or the other with respect to $p/\neg p$. Then, negating that will derive a weaker non-directional bias, namely that S *is* leaning one way or the other with respect to $p/\neg p$.

Finally, two reviewers ask how this relates to other accounts generally. I discuss a different version of my account relative to others in [4]. The present account represents progress over my account in [4] in that it provides a unified syntax/semantics for the left-periphery, with a clear dynamic view of context update, and a semantics for PPQs and HNQs that are directly related via entailment, which is a crucial precondition for the bias derivation in section 4. One advantage of this account is that the mechanics of the bias derivation is clear. Another is that the account does not unify HNQ bias with bias in other questions, such as polarity/verum focus questions, which appear to be empirically distinct (see discussion in [4]).

6 Conclusion

On the basis of a uniform commitment space semantics and syntax, the present account explains why preposed negation in polar questions conveys speaker bias crosslinguistically: because preposing *n't* scopes it over $\vdash$, which produces a less informative question than its competitor without *n't*.

[4] Let '*Bel*' stand for 'belief', i.e. doxastic necessity. Ignorance of whether p or $\neg p$ can be written as $\neg Bel\ p\ \wedge\ \neg Bel\ \neg p$. Then the negation of ignorance entails belief in either p or $\neg p$:
$\neg[\neg Bel\ p\ \wedge\ \neg Bel\ \neg p] = Bel\ p\ \vee\ Bel\ \neg p$

References

[1] Robert B. Brandom. Asserting. *Noûs*, 17:637–650, 1983.

[2] Daniel Büring and Christine A. Gunlogson. Aren't positive and negative polar questions the same? Ms. UCLA/UCSC, 2000.

[3] Kadir Gökgöz and Ronnie Wilbur. Positive bias in negative yes/no questions: Evidence for Neg-to-C in TİD. In Paweł Rutkowski, editor, *Different Faces of Sign Language Research*, volume 2, pages 15–42. The University of Warsaw, Faculty of Polish Studies, Warsaw, 2017.

[4] Daniel Goodhue. Isn't there more than one way to bias a polar question? *Natural Language Semantics*, pages 1–32, in press.

[5] Béata Gyuris. New perspectives on bias in polar questions: A study of Hungarian -*e*. *International Review of Pragmatics*, 9(1):1–50, 2017.

[6] Simone Hartung. Negation as a marker of commitment. UC San Diego, 2009.

[7] Manfred Krifka. Response particles as propositional anaphors. In *Proceedings of Semantics and Linguistic Theory (SALT)*, volume 23, pages 1–18, 2013.

[8] Manfred Krifka. Layers of assertive clauses: Propositions, judgments, commitments, acts. In Julia Hartmann and Angelika Wöllstein, editors, *Propositionale Argumente im Sprachvergleich: Theorie und Empirie*, pages 1–47. Gunter Narr, Tübingen, 2021.

[9] Manfred Krifka. Modelling questions in commitment spaces. In Moritz Cordes, editor, *Asking and Answering: Rivalling Approaches to Interrogative Methods.*, pages 63–95. Narr Francke Attempto, Tübingen, 2021.

[10] Manfred Krifka. Adjacency pairs in common ground update: Assertions, questions, greetings, offers, commands. In *SemDial 2022 (DubDial): The Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue*, pages 1–12, 2022.

[11] John MacFarlane. What is assertion? In Jessica Brown and Herman Cappelen, editors, *Assertion*. Oxford University Press, 2011.

[12] Charles Sanders Peirce. *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, MA, 1931–1958.

[13] Floris Roelofsen and Donka Farkas. Polarity particle responses as a window onto the interpretation of questions and assertions. *Language*, 91(2):359–414, June 2015.

[14] Maribel Romero and Chung-Hye Han. On negative *yes/no* questions. *Linguistics and Philosophy*, 27(5):609–658, 2004.

[15] Julian Schlöder and Alex Lascarides. Understanding focus: Pitch, placement, and coherence. *Semantics & Pragmatics*, 13(1):1–49, 2020.

[16] Robert Stalnaker. Assertion. In *Context and Content: Essays on intentionality in speech and thought*, pages 78–95. Oxford University Press, 1978.

[17] Tue Trinh. How to ask the obvious: A presuppositional account of evidential bias in english yes/no questions. In Luka Crnič and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, volume 2, pages 227–249. MITWPL 71, 2014.

[18] Robert van Rooij and Marie Šafářová. On polar questions. In *Semantics and Linguistic Theory*, pages 292–309, 2003.

When Maximality Is Not Maximally Informative: A Compositional Analysis of the Depth-Charge Illusion*

Vera Hohaus[1] and Nadine Bade[2]

[1] The University of Manchester, Manchester, U.K.
vera.hohaus@manchester.ac.uk
[2] Universität Potsdam, Potsdam, Germany
nadine.bade@uni-potsdam.de

Abstract

Building on existing proposals for the semantic analysis of negative antonyms [8, 19], intensional comparatives with *too* [11, 12], and monotonicity in comparison constructions [1, 2], we propose a compositional analysis of so-called depth-charge sentences like *No head injury is too trivial to be ignored*: The alleged illusion derives from the interaction of maximal informativity with the scalar properties of the degree set that underlies the standard of the comparison.

1 Introduction and Background

So-called depth-charge sentences like (1) have featured prominently in the psycholinguistics literature [3, 14, 16, 20] and are generally construed to entail (1-b): Regardless of their severity, all head injuries must receive medical attention. A large body of literature has taken this interpretation, while shown to be robustly available both in English and across languages [9, 13], to be an illusion that obscures the literal and compositional interpretation of the sentence in (1-a), under which all head injuries can or even should be ignored [4, 7, 22].

(1) *No head injury is too trivial to be ignored.* **"The Head Injury"**

a. %Ignore all head injuries!
b. ✓Do not ignore any head injuries!

Evidence for a view under which the prominent reading of (1) is an illusion comes from the structurally parallel case in (2), which is taken to convey that all missiles should be banned, as in (2-a), rather than legalised, as in (2-b).

(2) *No missile is too small to be banned.* **"The Missile Ban"**

a. ✓Ban all missiles!
b. #Do not ban any missiles!

However, with the exception of [14], no rigorous formal semantic analysis of the difference between the two examples has been attempted. The modest goal of this paper is to show that such an analysis is possible. We argue that the difference in interpretation is not an illusion and show that both readings can be derived compositionally. Building on existing formal analyses of negative-polar antonyms [19], modal comparison with *too* [11, 12], and monotonicity effects in the comparative constructions [1, 2], we propose that the alleged illusion derives from the interaction of maximal informativity with the scalar properties of the set of degree intervals that feeds into the standard of the intensional comparison.

*For helpful feedback and discussion, we would like to thank Dario Paape, Viola Schmitt, Ryan Walter Smith, Frank Sode, Siena Weingartz, Valerie Wurm, Malte Zimmermann, and Richard Zimmermann.

2 Towards the Analysis: Intensional Comparison and Negative Polarity

In this section, we review two of the key pieces of a compositional analysis of depth-charge sentences– the analysis of the syntax and semantics of intensional comparison with *too* [11, 12], and the semantic analysis of negative antonymity [8, 19].

Intensional Comparison. We take as our starting point one prominent analysis of *too* [11, 12] (but see also [5, 6, 21]), under which it encodes a greater-than comparison of two degree maxima, where the standard involves a covert possibility modal, as in (3). A sentence such as (4), under this analysis, has the Logical Form[1] sketched in (5) and is interpreted as in (6). Here, the comparison is between the maximum of the degree interval on the quality scale associated with the food F in the actual world $w_@$ and a degree interval whose right boundary corresponds to the highest quality at which it would be allowed to throw F away, a result of applying MAX to the set of degrees such there is a world in which F is thrown away; compare also Figure 1.

(3) $[\![\textit{too}]\!] = [\lambda Q \in D_{\langle\langle s,t\rangle,\langle s,t\rangle\rangle}. [\lambda R \in D_{\langle d,\langle s,t\rangle\rangle}. [\lambda w \in D_s.$ –to be revised–
$\text{MAX}(\lambda d.R(d)(w) = 1) > \text{MAX}(\lambda d*.Q(R(d*))(w) = 1)]]],$

where, for any set of degree intervals $P \in D_{\langle d,t\rangle}$,
$\text{MAX}(P) = \iota d\,[P(d) = 1 \;\&\; \forall d'\,[P(d') = 1 \rightarrow d \geq d']]$

(4) *The food is too good to be thrown away.*

(5) Logical Form for (4):

$[\, w_@\, [\, [\, \textit{the food}\,]\, [\lambda 1, e\, [\, [_{\text{DegP}}\, \textit{too}$
$[\, ①\, \text{(modal component)}\, \text{CAN}_{\langle\langle s,t\rangle,\langle\langle s,t\rangle,\langle s,t\rangle\rangle\rangle}\, [\, \text{PRO}_{1,e}\, \textit{be thrown away}\,]\,]$
$[\, ②\, \text{(degree set)}\, \lambda 2, d\, [\, t_{1,e}\, [_{\text{VP}}\, \textit{is}\, t_{2,d}\, \textit{good}_{\langle d,\langle e,\langle s,t\rangle\rangle\rangle}\,]\,]\,]\,]\,]\,]\,]\,]$

(6) Compositional Interpretation of (4),
where $w_@$ is the actual world and $[\![\textit{the food}]\!]^g =_{\text{simplified}} F$:

a. $[\lambda x : x \in D_e.\, [\![\textit{too}]\!]^g([\![①]\!]^g)([\![②]\!]^g)]\,([\![\textit{the food}]\!]^g)(w_@) = 1$
iff $\text{MAX}(\lambda d.\, \text{QUALITY}_{w_@}(F) \geq d) > \text{MAX}(\lambda d*.\, \exists w*\,[w*$ is accessible from $w_@$
$\&\; \text{QUALITY}_{w*}(F) \geq d* \;\&\; F$ is thrown away in $w*])$

b. $[\![①]\!]^g = [\lambda p \in D_{\langle s,t\rangle}.\, [\lambda w' \in D_s.\, \exists w*\,[w*$ is accessible from $w' \;\&\; p(w*) = 1$
$\&\; g(1,e)$ is thrown away in $w*]]]$

c. $[\![②]\!]^g = [\lambda d \in D_d.\, [\lambda w \in D_s.\, \text{QUALITY}_w(g(1,e)) \geq d]]$

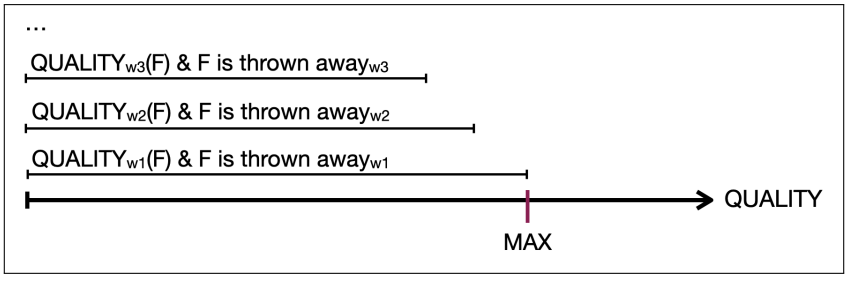


Figure 1: MAX $\gg \exists$ in the interpretation of the modal standard

[1] See also the more explicit tree structure for "The Missile Ban" in (9) below.

Negative Polarity. In the case of "The Head Injury" in (1) and "The Missile Ban" in (2), the comparison involves a negative antonym, however. Under an extent-based semantics [8, 19], these access the same scale as their positive-polar counterparts. In the antonym pair *old* and *young*, for instance, both adjectives relate entities to their extents on the age scale, as shown in Figure 2. The negative-polar *young* thus picks out those negative extents to which the entity is not old, as also indicated notationally by the $^-$ superscript.

(7) a. $[\![\, old \,]\!] = [\lambda d \in D_d \;\&\; d \in \text{AGE}^+ . [\lambda x \in D_e . [\lambda w \in D_s . \; \text{AGE}_w(x) \geq d]]]$
b. $[\![\, young \,]\!] = [\lambda d \in D_d \;\&\; d \in \text{AGE}^- . [\lambda x \in D_e . [\lambda w \in D_s . \; \text{AGE}_w(x) \geq d]]]$

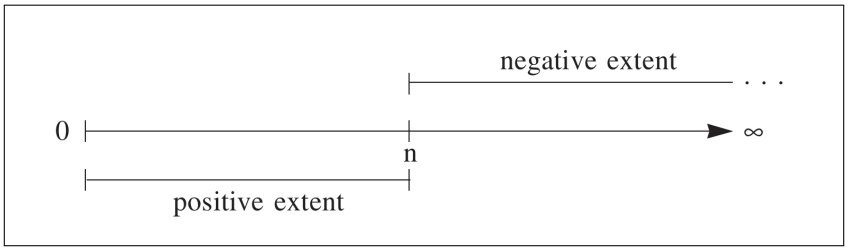


Figure 2: Negative extents [12]

A simple comparative like (8) with a negative-polar adjective will be true if and only if in the world of evaluation, the left boundary of Siena's longest negative age interval is lower on the age scale than the left boundary of Eve's longest negative age interval.[2] For an illustration of the mechanism behind determining the maximum of a negative extent, see also Figure 3.

(8) *Siena is younger than Eve.*
$\text{MAX}(\lambda d_1^- . \, \text{AGE}_{w_@}(S) \geq d_1) > \text{MAX}(\lambda d_2^- . \, \text{AGE}_{w_@}(E) \geq d_2)$
'The maximal age that Siena does not have
exceeds the maximal age that Eve does not have.'

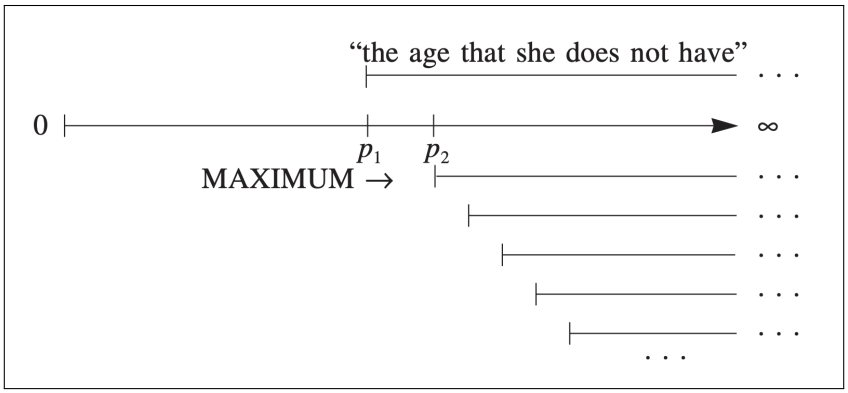


Figure 3: Maximal negative extents [12]

We put these two pieces (that is, the analysis of *too* as an intensional comparative, and the extent-based treatment of negative antonyms) together in the next section. We show that they derive the intended interpretation for "The Missile Ban", but only predict the largely unavailable reading for "The Head Injury", rather than the illusion. We then revise the analysis of *too* to incorporate insights from the literature [1, 2] on how to define maximality in the light

[2] Under a point-based view of degrees, this translates to Siena's age < Eve's age.

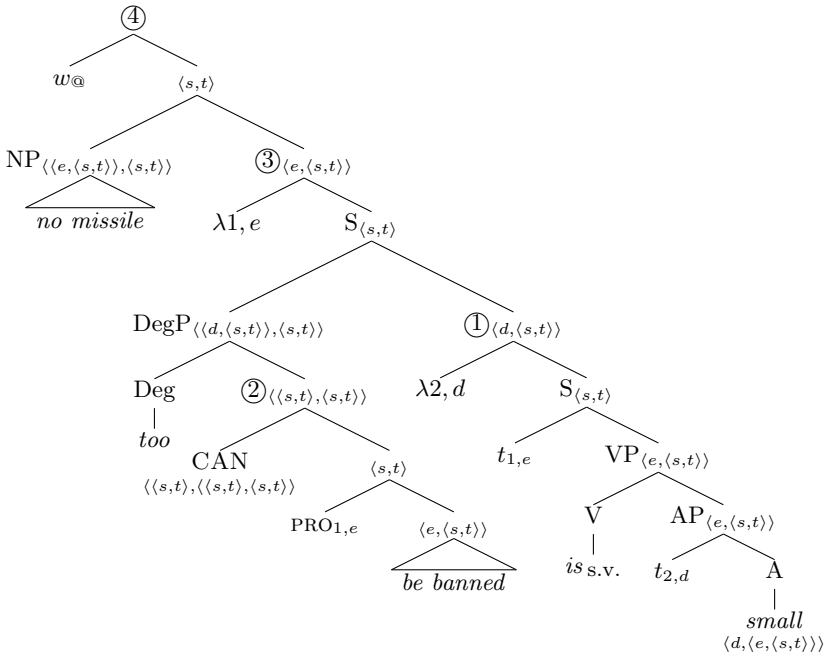
of the varying scalar properties of degree sets. We argue that the two key examples also vary in this respect, hence the different interpretations they make available.

3 Putting the Pieces Together, and a Revision

An analysis of "The Missile Ban" in (2) as a negative extent-based intensional comparison derives the observed interpretation, under which a ban is a possibility for all missiles, regardless of their size. We provide a Logical Form in (9) and sketch its interpretation in (10). The sentence is true if and only if, in the world of evaluation, there is no missile whose maximal negative size extent is greater than the standard of the comparison. The degree set whose maximal interval constitutes the standard of the comparison in this case is the set of negative size extent of $g(1, e)$ such that there is a world where $g(1, e)$ is banned.[3]

(2) *No missile is too small to be banned.*
(✓Ban all missiles! #Do not ban any missiles!)

(9) Logical Form for "The Missile Ban":



(10) Compositional Interpretation of "The Missile Ban":

a. $[\![①]\!]^g = [\lambda d : d \in D_d \;\&\; d \in \text{SIZE}^-. [\lambda w \in D_s. \text{SIZE}_w(g(1, e)) \geq d]]$

b. $[\![②]\!]^g = [\lambda p \in D_{\langle s,t\rangle}. [\lambda w' \in D_s. \exists w* [w* \text{ is accessible from } w' \;\&\; p(w*) = 1 \;\&\; g(1, e) \text{ is banned in } w*]]]$

c. $[\![③]\!]^g = [\lambda x \in D_e. [\lambda w \in D_s. \text{MAX}(\lambda d^-. \text{SIZE}_w(x)) > \text{MAX}(\lambda d*^-. \exists w* [w* \text{ is accessible from } w \;\&\; \text{SIZE}_{w*}(x) \geq d* \;\&\; x \text{ is banned in } w*])]]$

d. $[\![④]\!]^g = 1$ iff $\neg\exists x [x \text{ is a missile in } w_@ \;\&\; \text{MAX}(\lambda d^-. \text{SIZE}_{w_@}(x) \geq d) > \text{MAX}(\lambda d*^-. \exists w* [w* \text{ is accessible from } w_@ \;\&\; \text{SIZE}_{w*}(x) \geq d* \;\&\; x \text{ is banned in } w*])]$

[3] Under a point-based view of degrees, these truth conditions in (10) amount to requiring that there are no missiles whose size is smaller than the size from which onward a ban is a possibility.

For the structurally parallel "The Head Injury" in (1), this analysis however predicts only the reading whose availability is subject to variation, in (11), which may lead one to take the view that the other reading is indeed an illusion [14].

(1) *No head injury is too trivial to be ignored.*
(%Ignore all head injuries! ✓Do not ignore any head injuries!)

(11) Interpretation of "The Head Injury": – to be revised –
$\neg\exists x\,[x$ is a head injury in $w_@$ & $\textsc{max}(\lambda d^-.\,\textsc{severity}_{w_@}(x) \geq d)$
$> \textsc{max}(\lambda d*^-.\,\exists w*\,[w*$ is accessible from $w_@$ &
$\textsc{severity}_{w*}(x) \geq d*$ & x is ignored in $w*])]$

We suggest here that this conclusion is not inevitable: The maximality-based analysis of comparison that is also used for *too* has been challenged by examples like (12), which have a minimality, rather than maximality interpretation [1, 2]. They crucially also involve upward, rather than downward scalarity (that is, if one pound of flour is sufficient for this cake, two pounds will be as well, and so on). Adopting a definition of maximal informativity based on [1], in (13), derives the desired interpretation: The maximal informative amount of flour in these cases is the minimal amount of flour, not the maximal amount.

(12) *I have more flour than is sufficient to bake this cake.*
a. # I have more flour than the maximal amount that suffices to bake this cake.
b. ✓I have more flour than the minimal amount that suffices to bake this cake.

(13) For any $p \in D_{\langle d,\langle s,t\rangle\rangle}$ and world w, $\textsc{max}\text{-}\mathsf{inf}(p)(w) = \iota d\,[p(d)(w) = 1$ &
$\neg\exists d'\,[d \neq d'$ & $p(d')(w) = 1$ & [if $p(d')(w) = 1$, then $p(d)(w) = 1]]$

We follow the literature in assuming that the move away from maximality to maximal informativity is appropriate across the board, in comparison constructions [1, 2, 15], but also in the analysis of definite descriptions [17, 18], for instance. A revised lexical entry for *too* is in (14). Upon closer inspection, we find that the two key examples in (1) and (2) differ in the scalar properties of the degree set underlying the comparison; it is the informativity distinction between upward and downward scalarity which governs the available interpretations.

(14) $[\![\, too \,]\!] = [\lambda Q \in D_{\langle\langle s,t\rangle,\langle s,t\rangle\rangle}.\,[\lambda R \in D_{\langle d,\langle s,t\rangle\rangle}.\,[\lambda w \in D_s.$ – revised version –
$\textsc{max}\text{-}\mathsf{inf}(\lambda d.R(d)(w) = 1) > \textsc{max}\text{-}\mathsf{inf}(\lambda d*.Q(R(d*))(w) = 1)]]]$

To see this, consider Figure 4.



Figure 4: Upward/ downward scalarity and maximal informativity

While the negative size degrees under consideration for the comparison in "The Missile Ban" are those *from which onward* a ban is a possibility, this is not the case for "The Head Injury" under the default interpretation that has been characterised as an illusion. Here, the negative severity degrees under consideration in the standard of the comparison are those *up to which* such an

injury can be ignored. Crucially, given these differences, while in "The Missile Ban", MAX = MAX-inf, this is not the case for "The Head Injury", as shown in (15). Assuming downward monotonicity, the maximal informative negative extent is the minimal severity interval, whose lower bound is the degree up to which head injuries can be ignored (= the threshold from which treatment is required).[4]

(15) $\neg\exists x\,[x \text{ is a head injury in } w_@ \;\&\; \text{MAX-inf}(\lambda d^-.\,\text{SEVERITY}_{w_@}(x) \geq d) >$
$\text{MAX-inf}(\lambda d*^-.\,\exists w*\,[w* \text{ is accessible from } w_@ \;\&$
$\text{SEVERITY}_{w*}(x) \geq d* \;\&\; x \text{ is ignored in } w*])] =$

$\neg\exists x\,[x \text{ is a head injury in } w_@ \;\&\; \text{MAX-inf}(\lambda d^-.\,\text{SEVERITY}_{w_@}(x) \geq d) >$
$\text{MIN}(\lambda d*^-.\,\exists w*\,[w* \text{ is accessible from } w_@ \;\&$
$\text{SEVERITY}_{w*}(x) \geq d* \;\&\; x \text{ is ignored in } w*])]$

'No head injury has a negative severity extent that is greater than the minimal severity interval.'

The implausible (so-called literal) reading arises, we speculate, when the underlying degree set is re-conceptualised as upward-monotonous, that is, as having the monotonicity properties of the missile example. The negative severity degrees under consideration are then those *from which onward* no treatment is a possibility.

4 Some Concluding Remarks

Under the view presented in this paper, the depth-charge illusion is not an illusion, but rather just another case where maximality is not maximally informative. It is also another case where formal grammatical analysis can offer a systematic explanation of a phenomenon that has been labelled an illusion or fallacy [10], which highlights the importance of experimental research on meaning engaging with formal semantics research, and vice versa. The analysis presented above is compatible with many of the key observations in the literature, three of which we briefly mention here, even though we have to leave a more detailed discussion for another occasion. First, without world knowledge and lexical information, the monotonicity of the underlying degree set will end up being under-specified, and both interpretations should be equally plausible. For the example in (16), this indeed appears to be the case [20].

(16) *No wug is too dax to be zonged.*
(✓Do not zong any wugs! ✓Zong all wugs!)

Second, as it is informativity and its interaction with monotonicity that governs the available readings, similar contrasts should arise with intensional comparison with positive antonyms. Two relevant examples are in (17), which are interpreted, respectively, as encouragement to climb any mountain and as encouragement to not stop our efforts against polio [4].

(17) a. *No mountain is too high to climb.*
b. *No challenge is too big to stop us from saving our children from polio.*

Third, given the overall of the complexity of the key concepts involved (that is, degree comparison, negative antonymity, maximal informativity), the analysis presented here is compatible with the general observation that these sentences present a processing challenge, which may affect the availability of the different readings.

[4] Under a point-based view the interpretation in (15) corresponds to no head injury's severity being below the relevant threshold (indicated by MAX-inf in Figure 4).

References

[1] Sigrid Beck. Lucinda driving too fast again: The scalar properties of ambiguous *than*-clauses. *Journal of Semantics*, 30(1):1–63, 2013.

[2] Sigrid Beck and Hotze Rullmann. A flexible approach to exhaustivity in questions. *Natural Language Semantics*, 7(3):249–298, 1999.

[3] Paul Cook and Suzanne Stevenson. No sentence is too confusing to ignore. In *Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground*, pages 61–69, 2010.

[4] Egbert Fortuin. Deconstructing a verbal illusion: The 'no x is too y to z' construction and the rhetoric of negation. *Cognitive Linguistics*, 25(2):249–292, 2014.

[5] Valentine Hacquard. Aspect and actuality entailment: *Too* and *enough* constructions. In *Proceedings of Sinn und Bedeutung (SuB) 9*, pages 116–130, 2005.

[6] Valentine Hacquard. Aspects of *too* and *enough* constructions. In *Proceedings of Semantics and Linguistic Theory (SALT) 15*, pages 80–97, 2006.

[7] James Higginbotham. Is semantics necessary? In *Proceedings of the Aristotelian Society 88*, pages 219–241, 1988.

[8] Christopher Kennedy. *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*. PhD thesis, University of California, Santa Cruz, 1997.

[9] Johannes Kizach, Ken R. Christensen, and Ethan Weed. A verbal illusion: Now in three languages. *Journal of Psycholinguistic Research*, 45(3):753–768, 2016.

[10] Philipp Koralus and Salvador Mascarenhas. The erotetic theory of reasoning: Bridges between formal semantics and the psychology of deductive inference. *Philosophical Perspectives*, 27(1):312–365, 2013.

[11] Cécile Meier. *Konsekutive Konstruktionen und relative Modalität*. PhD thesis, Eberhard Karls Universität Tübingen, Tübingen, 2000.

[12] Cécile Meier. The meaning of *too*, *enough*, and *so... that*. *Natural Language Semantics*, 11(1):69–107, 2003.

[13] Dimitris Natsopoulos. A verbal illusion in two languages. *Journal of Psycholinguistic Research*, 14(4):385–397, 1985.

[14] Dario Paape, Shravan Vasishth, and Titus von der Malsburg. Quadruplex negatio invertit? The on-line processing of depth charge sentences. *Journal of Semantics*, 37(4):509–555, 2020.

[15] Jessica Rett. Antonymy in space and other strictly-ordered domains. *Baltic International Yearbook of Cognition, Logic and Communication*, 10:1–33, 2015.

[16] Anthony J. Sanford and Simon C. Garrod. The role of scenario mapping in text comprehension. *Discourse Processes*, 26(2-3):159–190, 1998.

[17] Philippe Schlenker. Informativity-based maximality conditions. *Snippets*, 26:18–19, 2012.

[18] Kai von Fintel and Sabine Iatridou. Definiteness as maximal informativeness. In Luka Crnič and Uli Sauerland, editors, *The Art and Craft of Semantics: A Festschrift for Irene Heim*, pages 165–174. MIT Working Papers in Linguistics, 2014.

[19] Arnim von Stechow. My reaction to Cresswell's, Hellan's, Hoeksema's and Seuren's comments. *Journal of Semantics*, 3(1-2):183–199, 1984.

[20] Peter C. Wason and Shuli S. Reich. A verbal illusion. *Quarterly Journal of Experimental Psychology*, 31(4):591–597, 1979.

[21] Linmin Zhang. *Enough*, *too*, and causal dependence. In *Proceedings of Sinn und Bedeutung (SuB) 22*, volume 2, pages 481–498, 2018.

[22] Thomas Ede Zimmermann and Wolfgang Sternefeld. *Introduction to Semantics: An Essential Guide to the Composition of Meaning*. De Gruyter, Berlin, 2013.

Compatibility, compossibility, and epistemic modality*

Wesley H. Holliday[1] and Matthew Mandelkern[2]

[1] University of California, Berkeley
wesholliday@berkeley.edu
[2] New York University
mandelkern@nyu.edu

Abstract

We give a theory of epistemic modals in the framework of possibility semantics and axiomatize the corresponding logic, arguing that it aptly characterizes the ways in which reasoning with epistemic modals does, and does not, diverge from classical modal logic.

1 Compatibility versus compossibility

The epistemic modals 'might' and 'must' have intriguing logical properties that fit poorly within standard approaches to modal logic. Most prominently, conjunctions of the form $\varphi \wedge \Diamond\neg\varphi$ or $\Diamond\neg\varphi \wedge \varphi$, dubbed *Wittgenstein sentences* after [24], appear to be contradictory. Evidence for this comes from the incoherence of sentences that embed Wittgenstein sentences, like (1)-(3):

(1) #Suppose John will win but he might not win. [25]

(2) #Either it's raining but it might not be, or it's not raining but it might be. [19]

(3) #Everyone who is a winner might be a loser. [10, 1, 26]

To bring out the significance of these patterns, compare (1)-(3) with minimal variants that embed Moore sentences [21] instead of Wittgenstein sentences:

(4) Suppose John will win but I don't know it.

(5) Either it's raining and I don't know it, or it's not raining and I don't know it.

(6) Everyone who is a winner is, for all I know, a loser.

These are, strikingly, coherent, unlike (1)–(3). While both Moore sentences and Wittgenstein sentences are unassertable, such contrasts in embedding behavior suggest that while Moore sentences are only pragmatically defective, Wittgenstein sentences are semantically defective.

The simplest way to diagnose this defectiveness is to say that Wittgenstein sentences are contradictions, that is, to adopt:

Wittgenstein's Law: $\varphi \wedge \Diamond\neg\varphi \vDash \bot$ and $\Diamond\neg\varphi \wedge \varphi \vDash \bot$.

However, in classical logic, Wittgenstein's Law leads to immediate trouble. If $\varphi \wedge \Diamond\neg\varphi \vDash \bot$, then it follows in classical logic that $\Diamond\neg\varphi \vDash \neg\varphi$ by the classical *Pseudocomplementation* rule:

Pseudocomplementation: If $\varphi \wedge \psi \vDash \bot$, then $\psi \vDash \neg\varphi$.

*This abstract presents ideas developed more fully in a corresponding paper [14]. For helpful feedback we thank three anonymous reviewers for the Amsterdam Colloquium, Maria Aloni, Alexandru Baltag, Justin Bledin, Kyle Blumberg, Sam Carter, Lucas Champollion, Ahmee Christensen, Yifeng Ding, Alexander Kocurek, J. Dmitri Gallow, Guillaume Massas, Eric Pacuit, Frank Veltman, and the audiences in the LIRa seminar at the University of Amsterdam in February 2022, the NASSLLI 2022 course on Possibility Semantics, and AiML 2022.

But, of course, $\Diamond\neg\varphi$ doesn't entail $\neg\varphi$: 'It might not be raining' doesn't entail 'It's not raining'.

We develop an approach to epistemic modals that adopts Wittgenstein's Law and (hence) rejects Pseudocomplementation. We do this by rejecting the semantic assumption which explains why Pseudocomplementation is classically valid: namely, that *compatibility entails compossibility*. In the classical possible worlds semantics for modal logic, whenever φ and ψ are compatible, i.e., neither entails the negation of the other, it follows that they are compossible, i.e., there is some world where they are both true and hence where $\varphi \wedge \psi$ is true. So, conversely, if $\varphi \wedge \psi$ is never true, it follows that φ and ψ are incompatible, and so ψ entails $\neg\varphi$.

However, we can block this reasoning, for we *need not* accept that compatibility entails compossibility. These are conceptually different notions: two sentences are compossible when they can both be true together; two sentences are compatible when neither entails the negation of the other. And, indeed, these notion can come apart in non-classical versions of a prominent alternative semantic framework for modal logic, namely *possibility semantics* [15, 12]. In possibility semantics, compatibility need not entail compossibility; and so, conversely, incompossibility need not entail incompatibility. The reason for this is that possibilities are *partial*, so that from φ not being true at a possibility, it does not follow that $\neg\varphi$ is true there. Instead, negation is interpreted using an independently specified relation of compatibility. So working with possibilities lets us develop a system where φ and $\Diamond\neg\varphi$ are incompossible—so that Wittgenstein's Law holds—while *also* maintaining that they are compatible—that is, that there are compatible possibilities x and y with φ true at x and $\Diamond\neg\varphi$ true at y, so that $\Diamond\neg\varphi \not\models \neg\varphi$.

2 Non-classicality (and its limits)

Before developing our system, we lay out some further desiderata. Pseudocomplementation is not the only classical rule that fails for epistemic modals, and we also want to capture some other failures. As [19] observes, *Distributivity* also appears to fail:

Distributivity: $\varphi \wedge (\psi \vee \chi) \dashv\models (\varphi \wedge \psi) \vee (\varphi \wedge \chi)$.

For if Distributivity were valid, then (7) would entail (2).

(7) It's raining or it's not raining, and it might be raining and it might not be raining.

But that is intuitively wrong: (7) is a coherent statement of ignorance, while (2) is incoherent.

Likewise, *Disjunctive Syllogism* appears to fail for epistemic modals [17]:

Disjunctive Syllogism: $(\varphi \vee \psi) \wedge \neg\psi \models \varphi$.

According to Disjunctive Syllogism, the entailment in (8) is valid, but that is intuitively wrong: the premise is just a statement of ignorance, while the conclusion is a claim about the world.

(8) Either Otto is inside or he must be outside, and he might be inside. So, Otto is inside.

We will take these apparent logical facts at face value by giving a theory on which Wittgenstein's Law holds and (hence) Pseudocomplementation fails (so that $\Diamond\neg\varphi \not\models \neg\varphi$), and Distributivity and Disjunctive Syllogism likewise fail. But, importantly, we aim to preserve as much as possible of the rest of classical modal logic. This last point is worth noting since some attempts to account for the non-classicality of epistemic modals lead to forms of non-classicality for which we think there is not clear evidence, as we discuss in § 6 below.

3 Possibilities

'Might' intuitively expresses *partiality of information*. While there are of course many ways to make sense of this in a framework whose indices are complete possible worlds, it is natural to model partiality of information with models whose indices themselves are partial: that is, which may leave unsettled the truth-value of some sentences. This is not a new idea: in particular, pioneering work in [22], and more recently [11, 2, 7], has given semantics for epistemic modals in terms of partial states. We will develop this idea in a different way, in a non-classical version (building on [8]) of *possibility semantics* for modal logic [15, 12]; in § 6 we will briefly compare this approach to other approaches that also build on partiality. (Possibilities are importantly different from truthmakers [6] but are arguably related to situations [18, 4].)

In possibility semantics, negation is definable in terms of a binary relation of *compatibility*: $\neg\varphi$ is true at possibility x just in case φ is not true at any possibility y compatible with x. This makes possibility semantics partial: it may be that φ is not true at x but is true at some y compatible with x, in which case $\neg\varphi$ is not true at x either. And this lets us draw a distinction that is impossible to draw in possible worlds semantics: we can say that two sentences φ and ψ are *incompossible*, in the sense that there are no possibilities where they are both true, while also maintaining that they are *compatible*, in the sense that there are compatible possibilities x and y with φ true at x and ψ true at y. Thus from $\varphi \wedge \psi \vDash \bot$ it does *not* follow that $\psi \vDash \neg\varphi$. We maintain that this is exactly the situation with φ and $\Diamond\neg\varphi$: these are never true together, but $\Diamond\neg\varphi \nvDash \neg\varphi$.

In more detail, we extend the possibility semantics for a propositional language from [8] to a modal language as follows. A *modal model* is a tuple $\mathcal{M} = \langle S, \between, V, R\rangle$ where S is a nonempty set (of possibilities); $\between$ is a reflexive, symmetric binary relation of *compatibility* on S; V is an atomic valuation, taking any sentence letter to a subset of S;[1] and R is a reflexive binary accessibility relation, representing epistemic possibility.[2] Given a modal model and possibility $x \in S$, we interpret our modal language with $[\![\cdot]\!]^x$, defined as follows:

- $[\![p]\!]^x = 1$ iff $x \in V(p)$, for sentence letters p;
- $[\![\varphi \wedge \psi]\!]^x = 1$ iff $[\![\varphi]\!]^x = 1$ and $[\![\psi]\!]^x = 1$;
- $[\![\Box\varphi]\!]^x = 1$ iff $\forall y \in R(x)\ [\![\varphi]\!]^y = 1$.

So far, $[\![\cdot]\!]^x$ is just as in possible worlds semantics for modal logic. But, crucially, we interpret negation using *incompatibility* rather than *lack of truth*:

- $[\![\neg\varphi]\!]^x = 1$ iff $\forall y \between x\ [\![\varphi]\!]^y \neq 1$.

Finally, we use De Morgan's law to define $\varphi \vee \psi$ as $\neg(\neg\varphi \wedge \neg\psi)$ and Duality to define $\Diamond\varphi$ as $\neg\Box\neg\varphi$, which yields:

- $[\![\varphi \vee \psi]\!]^x = 1$ iff $\forall y \between x\ \exists z \between y : [\![\varphi]\!]^z = 1$ or $[\![\psi]\!]^z = 1$.
- $[\![\Diamond\varphi]\!]^x = 1$ iff $\forall y \between x\ \exists z \in R(y)\ \exists u \between z : [\![\varphi]\!]^u = 1$.

[1] The subsets must meet a closure condition ensuring that propositions are well-behaved, or technically *regular*, where $A \subseteq S$ is regular iff $\forall x \in S$, if $x \notin A$, then $\exists y \between x\ \forall z \between y\ z \notin A$.

[2] R must also satisfy an interaction condition with $\between$ to ensure that modal sentences correspond to regular sets, namely: if $\exists z : xRz$ and $y \between z$, then $\exists x' \between x\ \forall x'' \between x'\ \exists z : x''Rz$ and $y \between z$.

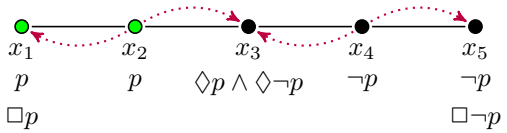


Figure 1: An epistemic model for a language with a sentence letter p; possibilities where p is true are in green. Solid lines represent $\between$ (compatibility), while dotted lines represent R (epistemic accessibility). Compatibility and epistemic accessibility are reflexive (reflexive arrows are omitted) but need not be transitive. Note the failure of Pseudocomplementation: at no possibility is $p \wedge \Diamond \neg p$ true, but $\Diamond \neg p$ does not entail $\neg p$, as the former is true at the middle possibility while the latter is not (neither p nor $\neg p$ is true there). Note also the failure of Distribution: $(p \vee \neg p) \wedge (\Diamond p \wedge \Diamond \neg p)$ is true at the middle possibility (as $p \vee \neg p$ is true at every possibility), while $(p \wedge \Diamond \neg p) \vee (\neg p \wedge \Diamond \neg p)$ is not true at any possibility.

The key move to enforce Wittgenstein's Law is to add a constraint stating that for any possibility x, there is a possibility y such that everything true at x *must* be true at y:

- *Knowability*: for all $x \in S$, there is a $y \in S$ such that for all $z \in R(y)$, $z \sqsubseteq x$,

where $z \sqsubseteq x$ ("z refines x") means that every possibility compatible with z is also compatible with x. That $z \sqsubseteq x$ ensures that every proposition true at x is also true at z. Hence *Knowability* says that every proposition true at x is true at all possibilities epistemically accessible from y and hence *must* be true at y. The intuition is that possibilities represent states that could be the information of some agent. So any possibility has to be "known" to obtain somewhere in modal space. Call a modal model that satisfies *Knowability* an *epistemic model*.

4 Consequence

We define consequence as usual: $\varphi \vDash \psi$ if at each point in each epistemic model, if φ is true, then ψ is true. This consequence relation has exactly the profile we set out to obtain.

First, Wittgenstein's Law is valid. Consider $\varphi \wedge \Diamond \neg \varphi$ (or $\Diamond \neg \varphi \wedge \varphi$, as the reasoning is the same). Suppose for contradiction this is true at x. By *Knowability*, there is a possibility y where this *must* be true, i.e., $[\![\Box(\varphi \wedge \Diamond \neg \varphi)]\!]^y = 1$. Hence by Distribution of $\Box$ over conjunction, $[\![\Box \varphi \wedge \Box \Diamond \neg \varphi]\!]^y = 1$. Then by Factivity for $\Box$, $[\![\Box \varphi \wedge \Diamond \neg \varphi]\!]^y = 1$. But by Duality and Double Negation Elimination (which is valid according to the semantics), $\Box \varphi \wedge \Diamond \neg \varphi$ is equivalent to the contradiction $\Box \varphi \wedge \neg \Box \varphi$ and so is never true. So Wittgenstein sentences are contradictory, since they are not "knowable" and so are not true at any possibility.

Importantly, not only are Wittgenstein sentences contradictions, but also they are *everywhere substitutable for contradictions*, meaning we accommodate incoherence data like (1)–(3) (once our system is extended with any reasonable treatment of attitudes and quantifiers). This is important to note, because in some systems Wittgenstein sentences are contradictions but cannot always be substituted for contradictions, as in domain semantics [25], which thus cannot account for the incoherence of disjunctive or quantified Wittgenstein sentences like (2) and (3).

Second, Pseudocomplementation fails, as desired: $\varphi \wedge \Diamond \neg \varphi$ is never true, but $\Diamond \neg \varphi \nvDash \neg \varphi$. This is because $\Diamond \neg \varphi$ and φ are still *compatible*, even though they are not compossible: there are possibilities x, y with $x \between y$ and $[\![\varphi]\!]^x = [\![\Diamond \neg \varphi]\!]^y = 1$, as in x_2 and x_3 in Figure 1 with p as φ. Such models show that $\Diamond \neg \varphi \nvDash \neg \varphi$, since $[\![\Diamond \neg \varphi]\!]^y = 1$ while $[\![\neg \varphi]\!]^y \neq 1$.

Finally, Distributivity and Disjunctive Syllogism are not valid. First, $(\varphi \vee \neg \varphi) \wedge (\Diamond \varphi \wedge \Diamond \neg \varphi)$ is consistent and equivalent to $\Diamond \varphi \wedge \Diamond \neg \varphi$, as desired. Contrary to Distributivity, it does *not*

entail $(\varphi \wedge \Diamond\neg\varphi) \vee (\neg\varphi \wedge \Diamond\varphi)$, which, by contrast, is a disjunction of contradictions and hence is contradictory. Second, $\varphi \vee \Box\neg\varphi$ is a tautology, so $(\varphi \vee \Box\neg\varphi) \wedge \Diamond\varphi$ is equivalent to $\Diamond\varphi$ and hence does not entail φ, contrary to Disjunctive Syllogism.

5 Logic

Principles 1-15 in Figure 2 constitute a logic that is sound and complete with respect to the possibility semantics of § 3. This logic extends *orthologic* (a weakening of classical logic that lacks Pseudocomplementation, Distributivity, and Disjunctive Syllogism [8]) with modal operators satisfying the logical analogue of *Knowability*, namely that if $\Box\varphi \vdash \bot$, then $\varphi \vdash \bot$.

It is worth noting that while this logic accounts for the non-classicality introduced by epistemic modals, it still preserves much of classical logic. Excluded Middle, Non-Contradiction, and De Morgan's laws remain valid. Even though it may be that neither φ nor $\neg\varphi$ is true at x, their disjunction will always be, given our clause for disjunction. $\varphi \wedge \neg\varphi$ cannot be true at any possibility since compatibility is reflexive. De Morgan's laws hold in virtue of our definition of disjunction and classical properties of our treatment of negation.

By extending our language with special Boolean propositional variables, standing for non-modal propositions, we can capture even more of classical logic. In [14], we further constrain our epistemic models so that we validate all classical reasoning in the Boolean fragment of our language (the closure of the Boolean propositional variables under $\wedge$ and $\neg$), as well as in fragments of the language that contain modals but do not mix different levels of modal embedding. Define *epistemic levels* inductively: the first epistemic level is the closure of the set of Boolean variables under $\wedge$ and $\neg$. For any n, construct the $n + 1^{st}$ epistemic level by taking the sentences with the form $\Box\varphi$, for φ in the n^{th} epistemic level, and closing under $\wedge$ and $\neg$. All the non-classicality introduced by epistemic modals comes from sentences that do not belong to any epistemic level (for instance, $\varphi \wedge \Diamond\neg\varphi$ conjoins a sentence φ of epistemic level n with a sentence $\Diamond\neg\varphi$ of epistemic level $n+1$). To capture this insight, we can further constrain epistemic models so that if φ and ψ are in the same epistemic level and are compatible in the model, then they are compossible in the model. Then the set of sentences at epistemic level n has a fully classical logic, as captured by principle 16 in Figure 2.

1. $\varphi \vdash \top$;	6. $\neg\neg\varphi \vdash \varphi$;
2. $\varphi \vdash \varphi$;	7. $\varphi \wedge \neg\varphi \vdash \psi$;
3. $\varphi \wedge \psi \vdash \varphi$;	8. if $\varphi \vdash \psi$ and $\psi \vdash \chi$, then $\varphi \vdash \chi$;
4. $\varphi \wedge \psi \vdash \psi$;	9. if $\varphi \vdash \psi$ and $\varphi \vdash \chi$, then $\varphi \vdash \psi \wedge \chi$;
5. $\varphi \vdash \neg\neg\varphi$;	10. if $\varphi \vdash \psi$, then $\neg\psi \vdash \neg\varphi$.
11. if $\varphi \vdash \psi$, then $\Box\varphi \vdash \Box\psi$;	14. $\Box\varphi \vdash \varphi$;
12. $\Box\varphi \wedge \Box\psi \vdash \Box(\varphi \wedge \psi)$;	15. if $\Box\varphi \vdash \bot$, then $\varphi \vdash \bot$;
13. $\varphi \vdash \Box\top$;	16. $\alpha \wedge (\beta \vee \gamma) \vdash (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$ provided α, β, γ are in the same epistemic level.

Figure 2: Principles 1-15 constitute the logic EO of *epistemic orthologic*, corresponding to epistemic models. Principles 1-10 characterize orthologics. Principle 15 is the proof-theoretic analogue of *Knowability*. Adding principle 16, corresponding to the extra constraint on epistemic models concerning epistemic levels in the main text, strengthens the logic from EO to EO^+.

6 Comparisons

While we do not have space for complete comparisons with other systems, we will briefly explain ways in which we think the approach outlined above improves on existing theories. The first point concerns non-classicality. While we can make sense of the failure of classical laws that indeed appear to fail for epistemic modals, we still preserve much of classical logic. This distinguishes our system from prominent systems like dynamic semantics and data semantics, where Excluded Middle fails [10, 22], and from state-based systems where De Morgan's laws fail [11].

A second point is methodological. Following the call in [13], we have given a sound and complete axiomatization of the validities of our system, thus directly characterizing the logical inferences that our system predicts will strike speakers as valid. Since capturing felt (in)validity is a central part of the task of natural language semantics, having such a characterization is an essential part of any adequate theory of a fragment of natural language.

The third point concerns empirical adequacy vis-à-vis Wittgenstein sentences. Unlike in dynamic semantics [23, 10], our account is order insensitive, capturing the infelicity of sentences that embed either $\varphi \wedge \Diamond\neg\varphi$ or $\Diamond\neg\varphi \wedge \varphi$. It is also fully general over substitution instances of φ; that is, it predicts Wittgenstein sentences are contradictions for any φ, unlike dynamic semantics, where this need not hold when φ is modal [20]. And Wittgenstein sentences are, again, not just contradictions but also everywhere substitutable for contradictions, giving our account considerably more empirical viability than systems like [25, 16, 3] where they are contradictions but not everywhere substitutable for contradictions.

Our final point starts with the observation that we can naturally form judgments about the probabilities of sentences containing modals and conditionals. In the full paper [14], we show how to add measures to possibility models to yield these judgments. Here we only make a conceptual point. Possibilities are (incomplete) ways for the world to be, and so it makes sense to specify measures over them. By contrast, in many treatments of epistemic modals that are likewise built around some notion of partiality, indices are treated as *information states* of some kind, and it is unclear how to construct measures over such things in a reasonable way (see [5, 9] for discussion). This problem is more tractable for some informational semantics than others. For instance, [9] shows how to extend probability measures to algebras of sets of pairs of a world and a set of possible worlds, which are the indices of domain semantics [25]. However, domain semantics is empirically untenable for the reasons noted above. The state-based semantics of [11] is empirically much more successful, but it faces a deep problem with probabilities, namely, that it yields $p \vDash \Box p$. This implies that any measure on an algebra of sets of states will always assign at least as great a probability to $\Box p$ as to p, but this is unrealistic: there is .5 probability that this fair coin will land heads but very little probability that it *must* land heads. By contrast, this problem does not arise for us, since we have $p \nvDash \Box p$.

7 Conclusion

To capture the way in which 'might' expresses partiality of information, we have developed a theory of epistemic modals built on possibility semantics, where indices are themselves partial. This lets us disentangle the notions of *compatibility* and *compossibility* that are run together in possible worlds semantics. Once we distinguish these, we can validate Wittgenstein's Law without arriving at the absurd conclusion that $\Diamond\neg\varphi \vDash \neg\varphi$. This approach captures the ways in which epistemic modals diverge from classical logic without introducing excessive non-classicality. In the full paper [14], we also give an algebraic characterization of our logic, equip our models with measures to capture probability judgments, and explore how to integrate conditionals.

References

[1] Maria Aloni. Conceptual covers in dynamic semantics. In Lawrence Cavedon, Patrick Blackburn, Nick Braisby, and Atsushi Shimojima, editors, *Logic, Language and Computation*, volume III. CSLI, 2000.

[2] Maria Aloni. FC disjunction in state-based semantics. Slides for Logical Aspects of Computational Linguistics (*LACL*), Nancy, France, 2018.

[3] Maria Aloni, Luca Incurvati, and Julian J. Schlöder. Epistemic modals in hypothetical reasoning. *Erkenntnis*, 2022.

[4] John Barwise and John Perry. *Situations and Attitudes*. MIT Press, 1983.

[5] Nate Charlow. Grading modal judgement. *Mind*, 129(515):769–807, 2019.

[6] Kit Fine. Truthmaker semantics. In Bob Hale, Crispin Wright, and Alexander Miller, editors, *Blackwell Companion to the Philosophy of Language*. Blackwell, 2017.

[7] Vera Flocke. Nested epistemic modals. Manuscript, Indiana University, Bloomington, July 2020.

[8] Robert I. Goldblatt. Semantic analysis of orthologic. *Journal of Philosophical Logic*, 3(1):19–35, 1974.

[9] Simon Goldstein. Epistemic modal credence. *Philosophers' Imprint*, 26, 2021.

[10] Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–216. Oxford, Blackwell, 1996.

[11] Peter Hawke and Shane Steinert-Threlkeld. Semantic expressivism for epistemic modals. *Linguistics and Philosophy*, 44:475–511, 2021.

[12] Wesley H. Holliday. Possibility semantics. In Melvin Fitting, editor, *Selected Topics from Contemporary Logics*, Landscapes in Logic, pages 363–476. College Publications, 2021.

[13] Wesley H. Holliday and Thomas F. Icard. Axiomatization in the meaning sciences. In Derek Ball and Brian Rabern, editors, *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, pages 73–97, Oxford, 2018. Oxford University Press.

[14] Wesley H. Holliday and Matthew Mandelkern. The orthologic of epistemic modals. Manuscript, arXiv:2203.02872, 2022.

[15] L. Humberstone. From worlds to possibilities. *Journal of Philosophical Logic*, 10(3):313–339, 1981.

[16] Luca Incurvati and Julian J. Schlöder. Epistemic multilateral logic. *Review of Symbolic Logic*, 2020.

[17] Nathan Klinedinst and Daniel Rothschild. Connectives without truth-tables. *Natural Language Semantics*, 20:137–175, 2012.

[18] Angelika Kratzer. An investigation of the lumps of thought. *Linguistics and Philosophy*, 12(5):607–653, 1989.

[19] Matthew Mandelkern. Bounded modality. *The Philosophical Review*, 128(1):1–61, 2019.

[20] Matthew Mandelkern. Dynamic non-classicality. *Australasian Journal of Philosophy*, 98(2):382–392, 2020.

[21] G. E. Moore. A reply to my critics. In P.A. Schilpp, editor, *The Philosophy of G.E. Moore*, volume 4 of *The Library of Living Philosophers*, pages 535–677. Northwestern University, 1942.

[22] Frank Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, 1985.

[23] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

[24] Ludwig Wittgenstein. *Philosophical Investigations*. Wiley-Blackwell, 2001, 3rd edition, 1953.

[25] Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

[26] Seth Yalcin. Epistemic modality *de re*. *Ergo*, 2(19):475–527, 2015.

THE HELL with questions*

Michela Ippolito

University of Toronto, Toronto, Ontario, Canada
michela.ippolito@utoronto.ca

Abstract

We discuss previous accounts of *wh-the-hell* questions, highlighting the challenges these accounts face and propose a novel analysis that has the potential to extend to *wh-the-hell* questions crosslinguistically.

1 The properties of *the hell* interrogatives

The hell questions have been argued to have syntactic and semantic properties that set them apart from canonical questions. In this section I introduce four properties that *the hell* have been claimed to have. The discussion in this paper will be limited to these properties.

First, *the hell* questions are **biased** and cannot be used in situations where an unbiased, neutral, information-seeking question is asked. In (1), since Zaira's question is meant to be a neutral information-seeking question, *the hell* is infelicitous.

(1) Zeno is telling her friend Zaira that yesterday he went to a new restaurant. Zaira is sincerely interested in what kind of food they prepare.
Zaira: #What the hell did you eat?

Intuitively and at a first approximation, a *the hell* question conveys that the speaker has a negative bias towards the answers to the question. For example, in (2), B's utterance conveys that they believe that they are not going to find people who trust XPharma anywhere (and that will volunteer for their medical trials).

(2) A: XPharma needs volunteers for their new experimental drug.
B: No way! Where the hell are they going to find people who trust them?

The negative bias, however, can coexist with the belief that there is a true answer to the question: in this case the question conveys speaker's ignorance and surprise. For example, in (3) the inspector knows that someone spoke to the reporters and the *the hell* question conveys that she is ignorant and utterly surprised that someone did.

(3) The inspector has just found out that confidential police information leaked to the press.
Inspector: Who the hell spoke to the reporters?

Second, *the hell* questions have been argued to be **aggressively non-D linked**, a property often viewed as connected to their incompatibility with the D-linked *wh* operator *which*, as shown in (4)

(4) *Which the hell spoke to the reporters?

*I would like to thank the audiences at different venues where this work was presented at different stages (ENS, Paris 8, University of Konstanz, IUSS, Università di Milano-Bicocca), as well as the *Amsterdam Colloquium 2022* reviewers, for helpful comments and suggestions. All mistakes are mine.

Third, *the hell* questions have a limited **embeddability**. They have been argued to be acceptable only under interrogative verbs like *wonder* ((5-a)), negative verbs like *refuse* ((5-b)), negative quantifiers like *nobody* ((5-c)), and factive verbs as long as they are negated ((5-d)-(5-e)).

(5) a. I wonder who the hell spoke to the reporters.
b. Zeno refused to tell me who the hell spoke to the reporters.
c. Nobody knows who the hell spoke to the reporters.
d. # I know who the hell spoke to the reporters.
e. I don't know who the hell spoke to the reporters.

In the course of the article, we will see that, despite what is often assumed in the literature, not all these properties correctly characterize *the hell* questions. [11] has challenged the idea that *the hell* questions must be non-D-linked and also shown that the equivalent of *the hell* in Japanese can combine with the Japanese equivalent of *which*. [10] has shown that, while there are constraints on the embeddability of *the hell*, these constraints are different from what was previously believed.[1]

1.1 Two previous proposals

1.1.1 *The hell* as a domain widening operator.

[4] proposes that a *wh-the-hell* (i) is a dependent polarity item, that is an expression with the semantics of an indefinite contributing a variable that cannot undergo existential closure at the top of the sentence; (ii) forces a domain extension; (iii) triggers a presupposition of negative attitude. Here we present challenges to (i) and (ii). With respect to (i), DG argue that this property explain why *the hell* can only be embedded in "non-veridical" contexts (e.g. negative verbs like *refuse*, negative quantifiers like *nobody*, and negated factive verbs): however, [10] correctly noticed that whether *the hell* can be embedded depends not on the veridical or non-veridical nature of the predicate (as DG would have it) but on whether the speaker is ignorant about the answer to the question: as long as the speaker is ignorant about the answer to the question, a *the hell* question is fine, as shown in (6).[2]

(6) a. He refused to tell me who the hell bought that book
b. #I refused to tell him who the hell bought that book.

As for (ii), D&G maintain that the domain of quantification of *the hell* questions is always extended to include both familiar and novel entities, and that this explains the non-D-linked requirement. However, *the hell wh* can be D-linked, *pace* [12]. In (7-a), a *the hell* question is perfectly felicitous even though the speaker is only considering the answers on the test ([11]).

(7) a. *During a multiple choice test*:
What the hell is the right answer?

The conclusion, thus, is that, since crucial aspects of DG's proposal (the polarity of *the hell wh*, the non-D-linking requirement, the domain widening) are problematic, it is untenable.

[1] As I said at the beginning of this section, the discussion here is limited to the aforementioned properties. For reasons of space, we cannot discuss other properties of *the hell* such as the incompatibility with **in-situ questions** and **sluicing**, as well as the unavailability of pair-list readings in multiple *wh* questions with *the hell*.

[2] Sincewhat is wrong with (6-b) is replated to the semantics of *the hell*, I replace the '*' in [10] with '#'.

1.1.2 *The hell* as a domain restrictor for the *wh*

[10] proposes that *the hell* restricts the domain of quantification of the *wh* operator. Following [9]'s work, Martin assumes that *wh* words have a focus value but not an ordinary value and that (as we are familiar with from the focus literature) the focus value is a set of alternatives. For example, the focus value of *what* is given in (8-a). The focus value of *the hell*, on the other hand, is the complement of the set of contextually salient entities (of the relevant type) C, i.e. the set of entities that are not contextually salient or "unfamiliar" ((8-b)). When the two combine, the result is the set of non-human entities that are not contextually salient ((8-c)).

(8) a. $[\![\text{what}]\!]^f = \{x_e : x \text{ is not human}\}$, ordinary value $[\![\text{what}]\!]^o$ undefined
b. $[\![\text{the hell}]\!]^f = \overline{C}$
c. $[\![\text{what the hell}]\!]^f = [\![\text{what}]\!]^f \cap [\![\text{the hell}]\!]^f = \{x_e : x \text{ is not human}\} \cap \overline{C}$

Martin's proposal is interesting but, like DG's proposal, it faces some challenges. The first problem is that Martin too builds the non-D-linked property in the semantics of *the hell* and, in light of examples like (7-a), this is problematic: in (7-a) the speaker is not asking which of the unfamiliar answers (other than the ones on the test) is the correct one. The same point can be made with *wh* questions followed by an alternative question where the alternatives are clearly and explicitly given by the alternative question.

(9) *Dante is having a hard time deciding what to order at a restaurant even though there are only two options.*
Beatrice: Come on! What the hell do you want? Pizza or pasta?

A related problem is that the incompatibility between *the hell* and *which* is also built in the semantics of *the hell*. Martin assumes that the focus value of a *which NP* phrase is the set of salient N, as shown in (4). When *which* and *the hell* combine, the result is the empty set.

(10) a. $[\![\text{which}]\!]^f = \lambda P_{<e,t>}.P \cap C$
b. $[\![\text{which the hell}]\!]^f = [\![\text{which}]\!]^f \cap [\![\text{the hell}]\!]^f = C \cap \overline{C} = \emptyset$

From a cross-linguistic perspective, this is problematic since there are languages that, unlike English, allow *which*-like operators to combine with *the hell*-like expressions. Japanese is one such language ([11]). Italian is another: *quale*, 'which', is compatible with *diavolo*, 'the hell'.

(11) Dante is taking forever to make his order. Beatrice says:
Andiamo! Hanno solo sushi, pasta e pizza. Quale diavolo vuoi?
Go-pres.1pl have-pres.3pl only sushi, pasta and pizza which devil want-pres.2sg
Come on! They only have sushi, pasta and pizza. What (lit: which) the hell to you want?

If applied to the crosslinguistic data, Martin's proposal incorrectly predicts that languages with *which*-like operators and *the hell* like phrases should not be able to combine them.

Lastly, the notion of "unfamiliarity" employed by Martin to refer to the focus value of *the hell* is unclear. In (12), *the hell* is felicitous even though the book Jack is reading is contextually salient, familiar to both discourse participants, and known to be such by both participants.

(12) *Jillian arrives at Jack's house and sees Jack reading a book that she told him not to read.*

Jillian: What the hell are you reading? I told you not to read that book!
Jack: Yes, you did, but I was curious!

The general conclusion is, while everyone agrees that there is something special about the meaning of THE HELL questions, neither type of proposal discussed above captures it successfully.

2 A novel proposal

I will make standard assumptions about the semantics of questions (cf. [6]). A constituent question of the form $wh_\alpha\ P$ denotes the set of propositions of the form $x\ P$, where $x \in D_\alpha$. Polar questions p? denote the set containing the sentence radical and its complement. This is shown in (13)

(13) a. $[\![\text{wh}_\alpha\ \text{P?}]\!]^C = \{p \in \wp(W) : p = P(x) | x \in C \subseteq D_\alpha\}$
b. $[\![\text{p?}]\!]^C = \{p, \overline{p}\}$

For reasons of space, I will not discuss issues of compositionality here. In very general terms, I assume that *the hell* is a *wh* modifier that syntactically attaches to the *wh* operator: semantically, it takes the *wh* term and a property P, which compose *via* point-wise functional application. I take this computation to happen within the focus dimension (cf. [9]). (14) shows the contribution of *the hell*:[3] it doesn't change the denotation of the *hell*-less question (as shown in (i)); it carries a presupposition (shown in (ii)); and it introduces a conventional implicature (shown in (iii)). K is a discourse structure ([5]). BEL_s, w is the set of x's beliefs in w.

(14) $$[\![\textit{the hell}\ Q_{wh}]\!]^{K,g,w} = \left\{ \begin{array}{l} (i)[\![Q_{wh}]\!]^{K,g,w} \\ (ii)\partial[\bigvee Q_{wh} \in T_K] \\ (iii)\forall p \in [\![Q_{wh}]\!]^{K,g,w}[\exists r \in \text{BEL}_{s,w} : p <_r \neg p] \end{array} \right\}$$

The presupposition in (ii) requires that the disjunction of the answers ($\bigvee Q$) be on Table_K (where Table_K is in K; cf. [5]). $\bigvee Q$ is equivalent to the proposition that there is a true answer to the question ([1]). The conventional implicature in (iii) represents the core meaning of *the hell* questions: for all propositions p in the denotation of the *wh*-question, there is some proposition r that the speaker believes such that, relative to r, p is less likely than $\neg p$; in other words, for each answer p, the speaker has some reason r to doubt p.[4] Note that the speaker might have other beliefs which are neutral with respect to whether p or $\overline{p}$ is true, or that might even make p more likely than $\overline{p}$.[5]

Summing up, the proposal is that *the hell* does not alter the denotation of the *the hell*-less question. It requires that the proposition that there is a true answer to the question be on the Table. Lastly, it conventionally implicates that the speaker has reasons to doubt all the

[3] I abstract away from the fact that *the hell* combines with *wh* first, and for convenience's sake I represent a *the hell* question as [*the hell* Q_{wh}].

[4] Saying that a belief r makes p more likely than $\overline{p}$ means that r makes the speaker's credence in p greater than the speaker's credence in $\overline{p}$.

[5] The role that r plays in (iii) is similar to the role that a conversational background plays in a modal statement: a conversational background is a set of propositions with respect to which a necessity or possibility statement is evaluated, and the latter may have a different truth value when evaluated relative to different conversational background (for example, simplifying the discussion greatly, in a world where people are required to be nice but they are known not to be, the necessity statement *Zeno must be nice* is deontically true, but epistemically false).

answers. In other words, on the basis of some salient facts, the speaker has a negative bias towards all individual answers being considered.

3 Explaining away our desiderata

No ∃-presupposition. *Contra* [14], a *the hell* question does not carry the existential presupposition that an answer to the question is true: $\bigvee Q$ is not required to be in the common ground but merely on the Table. This is good since in examples like (15), an existential presupposition would be too strong.

(15) a. What the hell can they possibly do at this point?
b. Who the hell is ever going to marry someone like that?

D-linked *the hell* questions are not problematic. *The hell* questions are just like any other question: if there are contextual restriction on the domain of the *wh*, the set of propositions in the denotation of the question will be similarly constrained. In the multiple choice example, repeated in (16), the answers being considered are the ones on the test: the conventional implicature is that for every proposition p of the form [x is the right answer], where x is one of the choices on the test, there is some proposition r that the speaker believes and relative to which, p is less likely than $\overline{p}$.

(16) *During a multiple choice test*:
What the hell is the right answer?

Note that this is compatible with the speaker believing that one of the answers on the test is the right answer. This creates doxastic tension but not a contradiction: on the one hand, the speaker has reasons to doubt each answer, but on the other hand she believes that one is true. Her body of beliefs is inadequate to arrive at the true answer, and the speaker realizes that. This analysis captures properties that seemed to be incompatible in other accounts: (i) the D-linked nature of the question; (ii) the belief that there is a true answer; (iii) the intuition that the speaker is clueless and that "no option is viable" ([13]).[6]

***The hell*-type phrases are semantically compatible with D-linked *wh* operators.** We saw above that our proposal does not involve any kind of domain widening or the selection of non-salient (unfamiliar) entities, therefore we predict no incompatibility between *which*-type operators and *the hell*-type phrases on the basis of their semantics. This correctly rules in the cross-linguistic data we discussed. As far as English in concerned, we tentatively conclude that what determines the incompatibility between *which* and *the hell* lies outside the semantics.

Requirement of speaker's ignorance explained. This proposal accounts for Martin's observation that speaker's ignorance is necessary to license *the hell*, as shown by (17).

(17) a. #I refused to tell him who the hell bought that book.
b. He refused to tell me who the hell bought that book.

[6] Let's consider briefly cases where there are only two alternatives in the focus value of the question: for example, our math test scenario where there are only two possible choices for each question. The speaker believes that one of the two answers (call them α and β) is correct (i.e. that there is a true answer to the question). The speaker may judge $\alpha <_r \overline{\alpha}$ with respect to some belief r, and she may judge $\beta <_{r'} \overline{\beta}$ with respect to some belief r'. If she also believes that there is a true answer to the question (a belief about tests in general, that one of the answers is the correct one), then she will be in what we called a state of doxastic tension, or a state of "unexpected ignorance", where her beliefs about the subject are at odds with some other belief she holds, and are therefore inadequate to find the true answer.

(17-a) is pragmatically infelicitous because there is a clash between the implication that the speaker knows who bought that book (their identity), and the conventional implicature of *the hell* that, the speaker has some reason to doubt each proposition of the form [x bought that book].[7]

Domain widening is an illusion. Domain widening is not part of the semantics (as we saw clearly with the D-linked cases), but at best a pragmatic inference that the domain of the question might not be adequate to provide a (positive) true answer.

4 Conclusion and open issues

While most of the examples of *the hell* questions we considered in the paper are biased questions, a *the hell* question can also be rhetorical: in (18) for example, the speaker is signalling to the addressee that nobody is ever going to marry Olivia, and that for the speaker the issue is closed, despite A's explicit assertion to the contrary.

(18) A: Someone will marry Olivia.
B: Who the hell will/would ever do that? That's nuts!

Following [7] and [8], I take rhetorical questions to be questions signalling not that the answer is in the common ground (*contra* [3], [2], a.o.), but that the speaker takes the issue raised by the question to be *closed*: this requirement can be satisfied in contexts where there is indeed doxastic *harmony* between speaker and addressee (and the issue is in the common ground), but also in contexts where there is doxastic *discord* between speaker and addressee but the speaker takes the issue to be closed, resolved in the discourse. Previous literature has only focussed on *harmony* cases (that is cases where the answer to the question is taken to be part of the common ground), thus missing the fact that a speaker can ask a rhetorical question even in the face of the fact that speaker and addressee explicitly disagree about the question.[8] Understood in the way just described, the analysis proposed here extends to *the hell* rhetorical questions as well.

We leave to another occasion, a more thorough presentation of our proposal and how it accounts for properties of *the hell* other than the ones discussed above (e.g. incompatibility with sluicing, in-situ, and pair-list reading). We also leave to the future a discussion of the connection between *the hell* questions (in English and other languages) and *hell*-less questions marked by a similar intonation as *the hell* questions and having seemingly the same meaning. Last we note a connection to be explored in the future between the semantics proposed here for *the hell* and the semantics proposed by [7] and [8] for the *mano a tulipano* (MAT) gesture in Italian.

[7] Note that this is different from the test example, where the speaker merely knows that *some* answer is correct. While the conventional implicature is the same, in (17-a), not only does the speaker conveys that she knows that someone bought that book, but that she knows the *identity* of the person x who bought that book, and at the same time she implicates that she has reason to doubt that x bought that book. (17-a) is infelicitous for the same reason why *#I know that Olivia bought that book but I have some reason to doubt that she did* is. Indeed, that a mere existential commitment is compatible with having reason to doubt each alternative is corroborated by the felicity of *I know that one of these two people bought that book but for each of them, I have some reason to doubt that they did.*

[8] Rhetorical questions of this kind are often characterized by a cluster of different markers, e.g. a non-neutral intonation, the use of NPI *ever*, the conjunction *and* at the beginning of the question (as in B's question: *A: Someone will help me*; *B: And who the hell will (ever) do that?*).

References

[1] Dorit Abusch. Presupposition triggering from alternatives. *Journal of Semantics*, 27(1):37–80, 2010.

[2] Maria Biezma and Kyle Rawlins. Rhetorical questions: Severing questioning from asking. In Dan Burgdorf, Jacob Collard, Sireemas Maspong, and Brynhildur Stefánsdóttir, editors, *Proceedings of Semantics and Linguistic Theory 27*, pages 302–322, 2017.

[3] Ivano Caponigro and Jon Sprouse. Rhetorical questions as questions. In Estela Puig-Waldmueller, editor, *Proceedings of Sinn und Bedeutung 11*, pages 121–133, 2007.

[4] Marcel den Dikken and Anastasia Giannakidou. From hell to polarity: "aggressively non-d-linked" wh-phrases as polarity items. *Linguistic Inquiry*, 22(1):31–61, 2002.

[5] Donka Farkas and Kim Bruce. On reacting to assertions and polar questions. *Journal of Semantics*, 27(1):81–118, 2010.

[6] C. L. Hamblin. Questions in montague english. *Foundations of Language*, 10:41–53, 1973.

[7] Michela Ippolito. Gestures as markers of non-canonical questions. In Julian J. Schlöder, Dean McHugh, and Floris Roelofsen, editors, *Proceedings of the 22nd Amsterdam Colloquium*, pages 574–583. University of Amsterdam, 2019.

[8] Michela Ippolito. The contribution of gestures to the semantics of non-canonical questions. *Journal of Semantics*, 38(3):363–392, 2021.

[9] Hadas Kotek. *Composing questions*. Linguistic Inquiry Monograph. MIT Press, 2018.

[10] Joshua Martin. Wh-the-hell as a polarity insensitive, speaker-oriented domain restrictor. In Joseph Rhyne, Kaelyn Lamp, Nicole Dreier, and Chloe Kwon, editors, *Proceedings from Semantics and Linguistic Theory 30*, pages 334–354. Linguistic Society of America, 2020.

[11] Takeshi Oguro. The true nature of wh-the-hell phrases. In Stefan Huber and Sonia Ramírez Wolmuth, editors, *2nd Annual Tampa Workshop pn Syntax, Semantics, and Phonology*, volume 4 of *Tampa Papers in Linguistics*. 2018.

[12] David Pesetsky. Wh-in-situ: movement and unselective binding. In Eric Reuland and A. ter Meulen, editors, *The representation of (in)definiteness*, pages 98–129. Cambridge, MA: MIT Press, 1987.

[13] Kyle Rawlins. *(Un)conditionals: An investigation in the syntax and semantics of conditional structures*. PhD thesis, UCSC, 2008.

[14] Anna Szabolcsi and Frans Zwarts. Weak islands and an algebraic semantics of scope taking. *Natural Language Semantics*, 1(3):235–284, 1993.

Why neg-raising requires stativity*

Paloma Jeretič[1] and Deniz Özyıldız[2]

[1]ZAS Berlin [2]Universität Konstanz

1 Introduction

The neg-raising (NR) inference is the inference from sentences like (1a) to ones like (1b), where negation is expressed on the matrix verb, but understood to negate the embedded proposition.

(1) a. Jo doesn't think that Su paid rent. b. Jo thinks that Su didn't pay rent.

The inference can be drawn with some predicates, like 'think,' 'believe' and 'want,' but not with others, like 'be certain,' 'claim' or 'wish.' A curious generalization, in (2), further links its availability to lexical aspect, and tells us what kinds of (uses of) predicates *cannot* trigger it (modified from Bervoets 2014: p. 112). Why this should hold is not yet well understood.[1]

(2) All NR predicates are stative, and in case an eventive counterpart of a NR predicate exists, this eventive counterpart does not trigger the NR inference.

In this paper, we focus on thought reports in French (which differentiates in the past between imperfective and perfective) and show that these are stative and NR in the imperfective, but eventive and non NR in the perfective—as expected per the generalization (Sect. 3). Adopting independently motivated assumptions about lexical and grammatical aspect (Sect. 2), we show that an analysis of the NR inference as a Scaleless Implicature (Jeretič 2021, 2022, see also Mirrazi and Zeijlstra 2021), but not as derived from an Excluded Middle presupposition (Bartsch 1973; Gajewski 2005), predicts this observation (Sect. 4). This result is a step towards understanding (2) in its full generality, as well as how logical and eventuality related properties of attitude predicates (like NR and aspect) interact.

2 The perfective and eventivity (background on aspect)

Much like their non attitude counterparts, attitude predicates can be categorized into ones that describe states ('know,' 'be certain'), and ones that describe events ('find out').[2] Often, we find that attitude predicates *alternate* between stative and eventive: In English, *think* may give rise to an ongoing state reading in the simple present, or to an ongoing event reading in the present

*Many thanks to Tanya Bondarenko, Aurore Gonzalez, Vincent Homer, Jérémy Pasquereau, and to our reviewers and audience. This work is partially supported by funding for the project "MECORE: Cross-linguistic investigation of meaning-driven combinatorial restrictions in clausal embedding" jointly awarded by the AHRC (AH/V002716) and DFG (RO 4247/5-1) to Romero, Uegaki and Roelofsen and the LeibnizDream project, awarded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (No 856421) to Sauerland, Alexiadou and Guasti.

[1]The first mention of this generalization that we know of is in Horn (1978: p. 206), who attributes it to Polly Jacobson. It seems to be found in early sources on NR under various guises, which oppose 'parenthetical,' 'performative' or 'metaphorical' uses to the 'literal' uses of a predicate, where only the former are NR (Horn and Bayer, 1984; Lakoff, 1969; Prince, 1976). The link with aspect receives renewed attention in work by Bervoets (2014, 2020), Xiang (2014), Özyıldız (2021) and Bondarenko (2022).

[2]More accurately, predicates taken together with their arguments and modifiers are categorized in this way.

progressive, in (3). Example (4) shows that a stative like *be mad* cannot occur in the simple present, and an eventive like *draw a circle*, in the progressive (Dowty 1979, a.o.).

(3) Calypso {thinks, is thinking} that it's raining.

(4) a. Zoe {is, #is being} mad. b. Zoe {#draws, is drawing} a circle.

These stative/eventive alternations are varied; for concreteness, we focus on the alternation between stative and eventive thought reports introduced by French *penser* (our findings extend to *croire*, 'believe'). We choose French because predicates may occur in the imperfective *imparfait* or the perfective *passé composé*, the latter forcing eventive interpretations.

(5) Hier, Calypso {pensait, a pensé} qu'il pleuvait.
Yesterday, Calypso thought$_{\{\text{IPFV, PFV}\}}$ that it was raining.

In general, both eventive and stative predicates, in (6a) and (7a), may occur in the imperfective, the former giving rise to an ongoing event, and the latter, to an ongoing state reading. In the perfective, however, while eventive predicates are acceptable, in (6b), statives are usually odd, in (7b). They become acceptable through contextual support, or with adverbs like *soudainement* ('suddenly') or *deux fois* ('two times'). When they are, the state that they describe is interpreted as bounded. We therefore rely on the perfective to identify bounded thought reports.[3]

(6) a. Hier, Zoé dessinait un cercle.
Yesterday, Zoé was drawing a circle.
b. Hier, Zoé a dessiné un cercle.
Yesterday, Zoé drew a circle.

(7) a. Hier, Zoé était en colère.
Yesterday, Zoé was$_{\text{IPFV}}$ mad.
b. Hier, Zoé a été en colère #(2 fois).
Yesterday, Zoé was$_{\text{PFV}}$ mad twice.

This restriction is at first sight unexpected. A common meaning for the perfective, attributed to Klein (1994), places the the runtime $\tau(e)$ of an event e within a contextually determined reference time R. Applied to *be mad*, we get the statement in (8), which should be satisfied by any state of Zoé being angry whose runtime is contained within yesterday. This is not odd in itself as some states, like being angry, hungry, etc., can be very temporary.

(8) $[\![\text{Zoé a été en colère}]\!] = \exists e.\tau(e) \subseteq \mathit{yesterday} \wedge \mathit{angry}(\mathit{zoé}, e)$

We follow Bary (2009) and Homer (2021) in assuming that statives are odd in the perfective because the perfective imposes a restriction on the eventuality descriptions it combines with, namely, that these be quantized. As defined in (9), a quantized predicate is one which, when it is true of an event, is false of any of its subevents. The perfective is defined, in (10a), with quantization as a definedness condition on its predicate argument, cf. the imperfective in (10b).

(9) $\text{QUA}(P)$ iff $\forall e \forall e'[(P(e) \wedge e' \sqsubset e) \rightarrow \neg P(e')]$

(10) a. $\text{PFV} := \lambda P : \text{QUA}(P).\exists e.P(e) \wedge \tau(e) \subseteq R$ b. $\text{IPFV} := \lambda P.\exists e.P(e) \wedge R \subseteq \tau(e)$

A predicate like *draw a circle* is quantized as no completed drawing of a circle is a proper subevent of another one, and hence, acceptable in the perfective. Stative predicates, however, are not quantized, and are divisive. This is defined in (11a): an eventuality true of a stative has all its subevents also true of that stative. To combine a stative with the perfective, then, one must first derive a quantized predicate from the stative. There are different ways of doing

[3] *Penser* has the option of being eventive in the imparfait as well. We have reason to think that these cannot be NR either, much like progressive thought reports in English (Özyıldız, 2021). We also leave the question of why bare *penser* sounds better in the perfective than bare *être en colère* to another occasion.

so, falling under the general notion of aspectual coercion (Moens and Steedman, 1988). We adopt Bary's MAX operator, defined in (11b), which takes a predicate and maximizes it, such that MAX(P) is true of P-events that have no proper P-subevents. Applied to a stative like 'be angry' or 'think p', it returns a predicate true of bounded events of being angry and thinking p. As such predicates are quantized, they are now free to combine with PFV as defined in (10a).

(11) a. P is stative iff $\forall e[P(e) \rightarrow \exists e'[e' \sqsubset e \wedge \forall e''[e'' \sqsubset e \rightarrow P(e'')]]]$
b. MAX $:= \lambda P \lambda e.P(e) \wedge \forall e'[e \sqsubset e' \rightarrow \neg P(e')]$ (Bary, 2009)

Implicit in this discussion is the assumption that *think* is stative at its core, and its eventive counterpart is derived from it via aspectual coercion. (There are alternatives, see Özyıldız 2021.)

Finally, it is crucial for us that negated statives involve aspectual coercion as well. One reason for thinking so is that the negation of a stative in the perfective remains odd, in (12).[4]

(12) Hier, Paul {n'était pas, #n'a pas été} en colère.
Yesterday, Paul was$_{\{\text{IPFV, PFV}\}}$n't mad.

3 NR requires stativity

Thought reports can be stative or eventive. If it is right that NR only arises with statives, we should find that stative thought reports may (and in fact tend to) give rise to the inference, but that their eventive counterparts may not. This is what we proceed to show now.[5]

The data supporting this result is delicate, and we first exercise our intuition on naturally occurring examples. Example (13a) is about carelessness in adopting children from areas struck by disaster, and (13b) reports on an unsuspecting shop owner, moments before she got robbed. Both examples feature perfective *penser* and can be translated naturally as "it did not occur to X that p," which does not involve NR. This is particularly clear in (13a), as it is surely false that the subjects have (had) the thought that adopted children would *not* grow up.

(13) a. On n'a pas pensé que les enfants adoptés deviendraient des adultes.
We didn't think$_{\text{PFV}}$ that adopted children would become adults.
b. Elle n'a pas pensé que c'était quelqu'un de mal intentionné et a ouvert.
She didn't think$_{\text{PFV}}$ that it was an ill-intentioned person, and opened the door.

We now turn to constructed examples, that show that at least in some contexts, NR is natural in the imperfective, but not in the perfective. First we use a continuation based test for NR inspired by Collins and Postal's (2014) "Do you agree?" test. The continuations in (14) attribute a positive thought to Isa's mother, which is recovered anaphorically. If the preceding utterance can be NR, it should be able to provide one and feel natural. In the absence of NR, the preceding utterance should entail the absence of a thought, and the continuation should feel unnatural, as it requires there to be one.

(14) Quand elle est entrée dans une église pour la première fois,
When she went into a church for the first time,

[4] Another one is that actuality entailments survive under negation: *Marie n'a pas pu dormir*, 'Marie was$_{\text{PFV}}$n't able to sleep,' entails Marie didn't sleep. According to Homer (2021), these arise because another ('actualistic') way of coercing statives into being quantized for the perfective (necessary in particular for certain modals). If the quantization requirement were lifted, we would not see actuality entailments under negation.

[5] To be clear, similar data is found in the references cited in the introduction. But the generalization is striking enough to warrant the discussion of additional corroborating—and indeed also limiting—examples.

a. Isa {pensait, a pensé} que Dieu n'existait pas. Sa mère pensait ça aussi.
Isa thought that God didn't exist. Her mom thought that as well.

b. Isa {ne pensait pas, #n'a pas pensé} que Dieu existait. Sa mère ça pensait aussi.
Isa didn't think that God existed. Her mom thought that as well.

c. Isa ne prétendait pas que Dieu existait. #Sa mère pensait/prétendait ça aussi.
Isa wasn't claiming that God existed. #Her mom thought/claimed the same.

In the control sentences with embedded negation, in (14a), the choice of the perfective vs. the imperfective does not affect the felicity of the continuation. But, when negation is pronounced on the matrix verb, in (14b), the continuation is acceptable with *penser* in the imperfective, but not in the perfective. This suggests that NR becomes unavailable then. We compare this state of affairs to the oddness that arises with a non NR predicate in the imperfective, in (14c).

Second, we test NR with strong NPIs. These are known to be licensed under local negation or negated NR predicates, but not under negated non-NR predicates (Horn 1978, a.o.). What we see in (15a) is that embedded negation can license the strong NPI *de la nuit* regardless of whether *penser* is in the perfective or the imperfective. With matrix negation, the NPI is licensed when the *penser* is in the imperfective, but not when it is in the perfective. Perfective *penser* patterns similarly to the non NR predicate *prétendre*, in (15c). This, again, is evidence that the perfective disrupts the availability of the NR inference.

(15) a. Ce matin, Jo {pensait, a pensé} que Al n'avait pas dormi de la nuit.
This morning, Jo thought that Al hadn't slept a wink.

b. Ce matin, Al {ne pensait pas, #n'a pas pensé} que Al avait dormi de la nuit.
This morning, Jo didn't think that Al had slept a wink.

c. Ce matin, Jo ne prétendait pas que Al avait dormi (#de la nuit).
This morning, Jo wasn't claiming that Al had slept a wink.

Despite the relative sharpness of these contrasts, we make a qualification to the generality of the claim that the perfective disrupts NR with thought reports—we believe it should be put on the record, so that we may return to it at another occasion. First, there are contexts in which negated perfective *think* is, or appears to be, NR. In (16), for instance, Jo is committed to the view that Al's work was not convincing. Note also that the continuation succeeds.

(16) Al: Alors, t'en as pensé quoi de mon travail?
So what did you think of my work?

Jo: J'avoue que j'ai pas pensé qu'il était très convaincant. Zoé a pensé ça aussi.
I admit that I didn't think that it was very convincing. Zoé thought that too.

Pending a more thorough empirical investigation, we mention a couple of ways of squaring this fact with the data in (13), (14) and (15), which suggested the absence of the inference in the perfective. First, and we thank Vincent Homer (p.c.) for pointing out this option to us, Jo's utterance could be euphemistic. Second, it could be that Al's question sets up the assumption that Jo is opinionated about whether the work was convincing (this is *Adressee Competence* associated with information seeking questions (Farkas, 2022)). And if one utters $\neg think\ p$ against the opinionated background $think\ p \vee think\ \neg p$, the stronger meaning $think\ \neg p$ is derived. This is Bartsch's (1973) pragmatic Excluded Middle presupposition, which we elaborate on in the next section. Another solution would be, if we are right in this paper's claim that MAX is the reason NR is disrupted, that there are other ways of making stative thought reports compatible with the perfective that do not disrupt NR.

Finally, we want to note that the generalization concerns belief predicates, but not all canonically NR predicates. In particular, at least the desire predicate *vouloir* ('want') triggers the inference as strongly in the perfective as it does in the imperfective.

(17) Ce matin, Axelle n'a pas voulu que Salomé dorme [de tout l'après-midi]$_{\text{NPI}}$.
This morning, Axelle didn't want$_{\text{PFV}}$ that Salomé sleep all afternoon.

Regarding this, there are reasons to think that NR arises from different mechanisms depending on the predicate, as discussed in Jeretič (2021), which makes it possible for the perfective to interfere with *think*'s NR inference, without interfering with *want*'s.

4 An explanation

***Think* as an eventuality predicate** To combine aspectual operators with *think*, we incorporate its Hintikkan (1969) definition into an eventuality predicate, as that is what these compose with. The function *dox* is then defined to return a set of worlds compatible with an agent's belief *states*, rather than their propositional beliefs (following Hacquard (2006), a.m.o.).

(18) a. $dox := \lambda x \lambda e.\{w \mid e \text{ is a belief state of } x \text{ and } w \text{ is compatible with the content of } e\}$
b. $[\![\textsf{think}]\!] := \lambda p \lambda x \lambda e.dox(x, e) \subseteq p$

Note that the output of *dox* when its eventuality argument is saturated by an event that is not a belief state e of x is the empty set. Before composing with aspect, then, "Zoé think that it's raining," is a predicate true of Zoé's belief states whose content entails that it is raining, or, those at which she believes this.

(19) $[\![\textsf{think}]\!]([\![\textsf{that it's raining}]\!])([\![\textsf{Zoé}]\!]) = \lambda e.dox(zoe, e) \subseteq rain$

How the Excluded Middle presupposition fares A prominent set of accounts of NR makes use of an Excluded Middle assumption.[6] We explore whether two of them—EM as a pragmatic (Bartsch, 1973) or a semantic (Gajewski, 2005) presupposition—can be extended to derive our facts. Bartsch seminally proposes that *think* triggers a pragmatic EM presupposition (PP). She formulates it, as is usual for attitude reports, in the simple present, as in (20).

(20) *PP:* Jiu thinks it's raining or Jiu thinks it's not raining.
PP & "Jiu doesn't think it's raining" ⇒ Jiu thinks it's not raining.

While she does not discuss cases where the thought report is in another tense or aspect, it is natural to imagine that a past perfective report, for example, also has its PP in the past perfective, in (21). Nothing then stops NR from being derived, just like it is in the present.

(21) *PP:* Jiu thought$_{\text{PFV}}$ it rained or Jiu thought$_{\text{PFV}}$ it didn't rain.
PP & "Jiu didn't think$_{\text{PFV}}$ it rained" ⇒ Jiu thought$_{\text{PFV}}$ it didn't rain.

This result is undesirable. One could say in response that the EM is not presupposed with perfective *think*. After all, there are contexts where NR is suspended. See (22), from Bartsch.

(22) Peter knows about Brutus and Caesar, but doesn't know if the two lived at the same time. He neither thinks that Brutus killed Caesar, nor that Brutus didn't kill Caesar.

[6] We do not discuss syntactic NR (Collins and Postal, 2014; Crowley, 2019), as it is unclear to us how different choices of matrix verb aspect could allow or block the syntactic raising of negation from an embedded clause.

When it is explicitly stated that the attitude holder is unopinionated, the EM cannot be presupposed, and NR is not derived. Perhaps, then, contexts in which perfective thought reports are uttered are (for some reason) typically ones in which the attitude holder is not unopinionated, hence ones in which NR is not derived. This is even consistent with our example (16), which could be argued to be NR despite the perfective because the context makes opinionatedness explicit. However, this pragmatic account is jeapordized by examples like (23) (and also (13)). Here, the speaker says that they did not (perfectively) think that the Earth was round.

(23) Ce matin, comme à peu près tous les matins, je n'ai pas pensé que la Terre était ronde.
This morning, like almost every morning, I didn't think$_{\text{PFV}}$ that the Earth was round.

This sentence is not NR, despite the fact that world knowledge dictates that the speaker must be opinionated about whether the Earth is round. This suggests that whatever is hindering NR in the perfective, it is not insufficient contextual support for the EM presupposition.

We now consider the more widely adopted view that EM is a *semantic* presupposition, as first proposed by Gajewski (2005), building on a suggestion by Heim (2000). As far as we know, there is no discussion of how the presupposition interacts with aspectual operators. We choose to recast the EM presupposition for *think* in an event-based framework as follows:

(24) $\mathsf{a\ think\ p} := \lambda e.dox(a,e) \subseteq p$ *presupposes:* $\forall e'[dox(a,e') \subseteq p \vee dox(a,e') \subseteq \neg p]$[7]

We will show that this presupposition predicts a contrast between imperfective and perfective, where NR is derived in the former but not the latter. Despite this welcome result, the result for the perfective is inaccurate, as it is different from a simple lack of NR. First, let us see how the presupposition in (24) correctly derives NR in the imperfective aspect.

(25) NEG [IPFV [a think p
ass: $\neg\exists e[R \subseteq \tau(e) \wedge dox(a,e) \subseteq p] \equiv \forall e[R \subseteq \tau(e) \rightarrow \neg(dox(a,e) \subseteq p)]$[8]
psp: $\forall e'[dox(a,e') \subseteq p \vee dox(a,e') \subseteq \neg p] \therefore \forall e[R \subseteq \tau(e) \rightarrow dox(a,e) \subseteq \neg p]$

A sentence like 'a didn't think$_{\text{IPFV}}$ p' as in (25) entails that all belief states of a whose runtime contains the reference time R have as content $\neg p$. If a has any R-containing belief state (which will generally be true), then it will have as content $\neg p$. This entails a NR inference.[9]

We now will see that in the perfective, the EM presupposition predicts disruption of NR, but not as desired. The disruptor is not perfective per se, but rather the quantization requirement that is necessary for any predicate to compose with the perfective, discussed in section 2.

(26) NEG [PFV [MAX [a think p
ass: $\neg\exists e[\tau(e) \subseteq R \wedge dox(a,e) \subseteq p \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]]$
$\equiv \forall e[\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)] \rightarrow \neg(dox(a,e) \subseteq p)]$
psp: $\forall e'[dox(a,e') \subseteq p \vee dox(a,e') \subseteq \neg p]$
$\therefore \forall e[\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow dox(a,e') \subseteq \neg p] \rightarrow dox(a,e) \subseteq \neg p]$

The MAX operator introduces a conjunct that contains the term $\neg(dox(a,e') \subseteq p)$. Together with the EM presupposition, this term is equivalent to $dox(a,e') \subseteq \neg p$. The result, instead of non-NR

[7] Another option we considered was the *existence* of an event of opinionatedness, whose runtime includes reference time. The predictions are almost identical, except if the domain of the subject's belief states is empty.

[8] Note that our semantics for the perfective and the imperfective under negation makes for truth conditions that are not restrictive enough. Our results should not be affected by this issue.

[9] This result is not exactly equivalent to the meaning of a sentence with negation in the lower clause, that would state the *existence* of a belief state of a of $\neg p$. Yet if we keep the events in our domain close to R, we get near equivalence, since if there is a belief state of $\neg p$ that contains R, all belief states that contain R are $\neg p$.

belief claim, is a tautology, given that 'think' is a stative predicate. It reads: every e running in R such that all superstates of e are $\neg p$-beliefs is also a $\neg p$-belief. This already follows from stativity (see (11a)) of *think*: e, as a substate of a $\neg p$-belief state, is also a $\neg p$-belief state.[10]

Furthermore, we note that we also get an unexpected result in a positive quantized *think* utterance. In this case, the term $\neg(dox(a, e') \subseteq p)$ is still present, and again neg-raised. The result in (27) is a contradiction: it states that there is a p-belief state of a whose superevents are $\neg p$-belief states of a. A state that makes P true has all of its subevents make P true. Assuming that one cannot hold a p belief and a $\neg p$ belief at once, this is a contradiction.

(27) PFV [MAX [a think p
ass: $\exists e[\tau(e) \subseteq R \wedge dox(a, e) \subseteq p \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a, e') \subseteq p)]]$
psp: $\forall e'[dox(a, e') \subseteq p \vee dox(a, e') \subseteq \neg p]$
$\therefore \exists e[\tau(e) \subseteq R \wedge dox(a, e) \subseteq p \wedge \forall e'[e \sqsubset e' \rightarrow dox(a, e') \subseteq \neg p]]$

A different solution could say local accommodation of the EM presupposition—but then we'd have to say that it is systematically accommodated in the perfective. Why would that be?

Deriving appropriate NR disruption with scaleless implicatures In the previous section, we saw that the quantization conjunct is a source of disruption of NR in the perfective aspect. This will prove to be again the case if we take NR to be a scaleless implicature, but this time with a correct result, namely a plain meaning for negated perfective *think* with no NR.

In Jeretič (2022), following Jeretič (2021), NR is analyzed as a scaleless implicature, i.e. a strengthening from the reading given by the LF where *think* scopes below negation to the stronger NR reading. Scaleless implicatures are predicted by theories of scalar implicatures derived in the grammar, such as Bar-Lev and Fox (2020), whenever a quantifier has subdomain alternatives but no scalar alternative. In a nutshell, NR is the result of an exhaustivity operator EXH, as schematized below (aspectual operators ignored).

(28) $[\![$ EXH [NEG [a think p]] $]\!] \equiv [\![$ a think NEG p $]\!]$

The existence of scalar alternatives to a quantifier are determined based on the available lexemes in a language. In English and French, the word for 'think' can qualify as lacking a scalar alternative, because it makes a universal claim about the doxastic set of an individual, and has no obvious existential scalemate in the lexicon. Subdomain alternatives, in contrast, are lexically specified. Typically, the definition of a subdomain alternative is an expression obtained by replacing the domain of a quantifier with a non-empty subset; however for *think*, the definition is more complicated, because its domain contains bound variables. For this, we follow Jeretič (2022) in arriving at subsets of the assignment-dependent domain via choice functions which pick out members of $\mathcal{P}(dox(x, e))/\varnothing$ (the set of non-empty subsets of $dox(x, e)$), for each possible assignment of x and e. Thus, the set of subdomain alternatives for *think* correspond to replacing $dox(x, e)$ with $F(\mathcal{P}(dox(x, e))/\varnothing)$ for all possible choice functions F.

(29) $\text{Alt(think)} = \{\lambda p.\lambda x \in D_e.\lambda e \in E_v.F(\mathcal{P}(dox(x, e))/\varnothing) \subseteq p \mid F$ is a choice function defined on $\mathcal{P}(dox(x, e))/\varnothing$ for all $e \in E_v$ and $x \in D_e\}$ (Jeretič, 2022)

[10] Perhaps we were wrong about our non-trivial assumption that EM affects the MAX conjunct. Indeed, we could have the presupposition quantify over the domain of events D shared with that of the main assertion, but not D', that of the quantized conjunct (the assertion remains consistent if we can have $D \subset D'$). However doing so derives the unattested meaning that either all belief states of a in R are $\neg p$, or there is a non-quantized p-belief state in R, shown here: $\forall e \in D[\tau(e) \subseteq R \wedge \forall e' \in D'[e \sqsubset e' \rightarrow \neg(dox(a, e') \subseteq p)] \rightarrow dox(a, e) \subseteq \neg p]$.

Following Bar-Lev and Fox (2020), implicatures are derived by applying an operator EXH to a sentence S and its alternative set Alt(S). EXH asserts S, negates all Innocently Excludable (IE) alternatives (whose negation can be conjoined non-arbitrarily and consistently to S) and asserts all Innocently Includable (II) ones (which can be conjoined non-arbitrarily and consistently to IE-strengthened S). See formal definition in the Appendix.

We first show how NR is derived when EXH is applied to negated imperfective *think*. We begin with the LF in (30) before exhaustification, with its translation in (30a) and its set of alternatives in (30b). (31) is the result of applying EXH to (30) (see Appendix for full derivation).

(30) NEG [IPFV [a think p

a. **ass:** $\neg\exists e[dox(a,e) \subseteq p \wedge R \subseteq \tau(e)]$

b. **alts:** $\{\neg\exists e[F(\mathcal{P}(dox(a,e))/\varnothing) \subseteq p \wedge R \subseteq \tau(e)] \mid F$ is a choice function defined on $\mathcal{P}(dox(x,e))/\varnothing$ for all $e \in E$ and $x \in D\}$

(31) $[\![$ EXH [(30)][Alt((30))] $]\!] \equiv \neg\exists e[dox(a,e) \subseteq p \wedge R \subseteq \tau(e)] \wedge$ $\boxed{\forall e[R \subseteq \tau(e) \rightarrow dox(a,e) \subseteq \neg p]}$

The derived inference states that all belief states of a running through R are $\neg p$-beliefs, which entails the NR inference. (Note we have the same result with the EM presupposition in (19).)

We now show the failure to derive NR with the perfective. (32) is the LF of a perfective *think* utterance before EXH applies. In (32a) is its translation, and in (32b) its alternatives.

(32) NEG [PFV [MAX [a think p

a. **ass:** $\neg\exists e[\tau(e) \subseteq R \wedge dox(a,e) \subseteq p \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]]$
$\equiv \forall e[\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)] \rightarrow \neg(dox(a,e) \subseteq p)]$

b. **alts:** $\{\neg\exists e[F(\mathcal{P}(dox(a,e))) \subseteq p \wedge \tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg F(\mathcal{P}(dox(a,e'))) \subseteq p)]] \mid F$ is a choice function defined on $\mathcal{P}(dox(a,e))/\varnothing$ for all $e \in E\}$

The crucial detail in (32b) is that a subdomain alternative is defined by replacing *each* instance of *dox* with a given subset. This follows if we assume that the alternatives of an expression are obtained by pointwise function application from the alternatives of its parts. Thus, if MAX combines with P whose alternative set is Alt(P), the alternative set of MAX(P) will be obtained by slotting each alternative of P into each instance of the lambda-bound P in the expression for MAX. Recall that MAX, defined in (11b), selects for a predicate P and makes it appear twice in its assertion. This means that a given alternative of MAX(a think p) is obtained by having substituted an alternative of a think p into both of MAX's P-slots.

As also seen in the EM section, at these two places we have negated p-beliefs, which are thus candidates for NR. And indeed, EXH will locally derive NR at both places, as seen in (33).

(33) $[\![$EXH[(32)][Alt((32))]$]\!] \equiv \neg\exists e[dox(a,e) \subseteq p \wedge \tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]]$
$\wedge$ $\boxed{\forall e[(\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow dox(a,e') \subseteq \neg p]) \rightarrow dox(a,e) \subseteq \neg p]}$

The complete resulting inference, instead of being NR, is a tautology, in the same way that the result with the EM presupposition in (26) is tautological (so we do not repeat the explanation here). However the difference between this result and the one obtained with the EM presupposition is that on the presuppositional view, the assertion itself together with the presupposition *is* a tautology, and the non-NR meaning is unrecoverable. Here on the other hand, we have the base negated perfective meaning left intact, and it is simply *conjoined* to the tautological inference (as one does with implicatures). We therefore are left with an expression equivalent to the unenriched meaning of a negated perfective thought report, as desired.

References

Bar-Lev, M. E. and Fox, D. (2020). Free choice, simplification, and innocent inclusion. *Natural Language Semantics*, 28(3):1–49.

Bartsch, R. (1973). ,,Negative Transportation" gibt es nicht. *Linguistische Berichte*, 27(7):1–7.

Bary, C. (2009). *Aspect in Ancient Greek: A semantic analysis of the aorist and imperfective*. PhD thesis, Radboud Universiteit Nijmegen.

Bervoets, M. (2020). *The Semantics of Opinion*. Springer.

Bervoets, M. J. (2014). *On Opining: Modal Verbs, Dispositions, Free Choice, and Negation*. PhD thesis, UCLA.

Bondarenko, T. (2022). *Anatomy of an attitude*. PhD thesis, MIT.

Collins, C. and Postal, P. M. (2014). *Classical NEG raising: An essay on the syntax of negation*, volume 67. MIT Press.

Crowley, P. (2019). Neg-raising and neg movement. *Natural Language Semantics*, 27(1):1–17.

Dowty, D. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative seman tics and in Montague's PTQ*. Springer Science & Business Media.

Farkas, D. F. (2022). Non-Intrusive Questions as a Special Type of Non-Canonical Questions. *Journal of Semantics*, 39(2):295–337.

Gajewski, J. (2005). *Neg-raising: polarity and presupposition*. PhD thesis, MIT.

Hacquard, V. (2006). *Aspects of modality*. PhD thesis, MIT.

Heim, I. (2000). Degree operators and scope. In Jackson, B. and Matthews, T., editors, *Proceedings of SALT 10*, pages 40–64.

Hintikka, J. (1969). Semantics for propositional attitudes. In *Philosophical Logic*. D. Reidel, Dordrecht.

Homer, V. (2021). Actualistic Interpretations in French. *Semantics and Pragmatics*, 14(2).

Horn, L. (1978). Remarks on neg-raising. In Cole, P., editor, *Syntax and Semantics 9: Pragmatics*, pages 129–220. New York, NY: Academic Press.

Horn, L. and Bayer, S. (1984). Short Circuited Implicature: A Negative Contribution. *Linguistics and Philosophy*, 7(4):397–414.

Jeretič, P. (2021). *Neg-raising modals and Scaleless Implicatures*. PhD thesis, New York University.

Jeretič, P. (2022). Neg-raising as a scaleless implicature. To appear in *Proceedings of SALT 32*.

Klein, W. (1994). *Time in Language*. Routledge: London and New York.

Lakoff, R. (1969). A syntactic argument for negative transportation. In *Papers from the 5th Annual Meeting, Chicago Linguistics Society*, pages 140–147.

Mirrazi, Z. and Zeijlstra, H. (2021). Neg-raising without lexical excluded middle inferences: Resolving Horn Clauses and other problems. In *Proceedings of the 23rd Seoul International Conference on Generative Grammar (SICOGG 23): A Comparative Approach to the Syntax-Semantics Interface: 299.*, page 299.

Moens, M. and Steedman, M. (1988). Temporal ontology and temporal reference. *Computational linguistics*, 14(2):15–28.

Özyıldız, D. (2021). *The event structure of attitudes*. PhD thesis, UMass Amherst.

Prince, E. (1976). The Syntax and Semantics of Neg-Raising, with Evidence from French. *Language*, 52(2):404–426.

Xiang, Y. (2014). Obligatory Neg-raising in Mandarin: Negatives and Aspects. Unpublished manuscript, Harvard University.

Appendix

Definition of exh. We define our exhaustivity operator below, taken from Bar-Lev and Fox (2020).

(34) a. $IE(p, C) = \bigcap\{C' \subseteq C : C' \text{ is maximal \& } \{\neg q : q \in C'\} \cup \{p\} \text{ is consistent}\}$
b. $II(p, C) = \bigcap\{C'' \subseteq C : C'' \text{ is maximal \& } \{r : r \in C''\} \cup \{p\} \cup \{\neg q : q \in IE(p, C)\}$ is consistent$\}$

(35) $[\![\text{EXH}]\!](C)(p)(w) \equiv \forall q \in IE(p, C)[\neg q(w)] \wedge \forall r \in II(p, C)[r(w)]$

NR with negated imperfective thought reports.

(36) NEG [IPFV [a think p
a. **ass:** $\neg\exists e[R \subseteq \tau(e) \wedge dox(a, e) \subseteq p]$
$\equiv \forall e[R \subseteq \tau(e) \rightarrow \neg(dox(a, e) \subseteq p)]$
b. **psp:** $\forall e'[dox(a, e') \subseteq p \vee dox(a, e') \subseteq \neg p]$
c. $\therefore \forall e[R \subseteq \tau(e) \rightarrow dox(a, e) \subseteq \neg p]$

To show the derivation, we use an unrealistic model containing two eventualities associated with 2-world doxastic states: $dox(a, e_1) = \{w_1, w_2\}$, $dox(a, e_2) = \{w_3, w_4\}$. An example alternative of (36) evaluated with respect to this toy model is shown in (37).

(37) $\neg(\{w_1, w_2\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_3\} \subseteq p \wedge R \subseteq \tau(e_2))$

There are no IE alternatives: we can see this by attempting to exclude the strongest alternatives, i.e. those least likely to affect the result, which correspond to those obtained by replacing the domain with a singleton subset. The conjunction of the negation of these alternatives is equivalent to the negation of the prejacent. Excluding them would result in a contradiction, and excluding some but not all of these would be arbitrary; this means there are no IE alternatives.

(38) $[\{w_1\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_3\} \subseteq p \wedge R \subseteq \tau(e_2)]$
$\wedge[\{w_1\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_4\} \subseteq p \wedge R \subseteq \tau(e_2)]$
$\wedge[\{w_2\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_3\} \subseteq p \wedge R \subseteq \tau(e_2)]$
$\wedge[\{w_2\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_4\} \subseteq p \wedge R \subseteq \tau(e_2)]$
$\equiv \exists e[dox(a, e) \subseteq p \wedge R \subseteq \tau(e)]$

In contrast, all alternatives are II; we show in (39) the result of conjoining the same singleton-based alternatives to (36) (the non-singleton-based alternatives are also II; but their inclusion doesn't further affect the result, as we have already obtained a universal claim, the strongest result possible).

(39) $[\![\text{EXH } [(36)][\text{Alt}((36))]]\!] \equiv \neg\exists e[dox(a,e) \subseteq p \wedge R \subseteq \tau(e)]$
$\wedge\neg(\{w_1\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_3\} \subseteq p \wedge R \subseteq \tau(e_2))$
$\wedge\neg(\{w_1\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_4\} \subseteq p \wedge R \subseteq \tau(e_2))$
$\wedge\neg(\{w_2\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_3\} \subseteq p \wedge R \subseteq \tau(e_2))$
$\wedge\neg(\{w_2\} \subseteq p \wedge R \subseteq \tau(e_1) \vee \{w_4\} \subseteq p \wedge R \subseteq \tau(e_2))$
$\equiv \neg\exists e[dox(a,e) \subseteq p \wedge R \subseteq \tau(e)] \wedge \forall e[R \subseteq \tau(e) \rightarrow dox(a,e) \subseteq \neg p]$

No NR with negated perfective thought reports.

(40) NEG [PFV [MAX [a think p

a. **ass:** $\neg\exists e[\tau(e) \subseteq R \wedge dox(a,e) \subseteq p \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]$
$\equiv \forall e[\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)] \rightarrow \neg(dox(a,e) \subseteq p)]]$

b. **alts:** $\{\neg\exists e[F(\mathcal{P}(dox(a,e))) \subseteq p \wedge \tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg F(\mathcal{P}(dox(a,e'))) \subseteq p)]] \mid F$ is a choice function defined on $\mathcal{P}(dox(a,e))/\varnothing$ for all $e \in E\}$

Like above, no alternative is IE (left to the reader), but all alternatives are II. The effect of EXH is in (41); we show the inclusion of one alternative, the rest to be filled out.

(41) $[\![\text{EXH}[(40)][\text{Alt}((40))]]\!] \equiv \neg\exists e[dox(a,e) \subseteq p \wedge \tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]]$
$\wedge\neg(\{w_1\} \subseteq p \wedge \tau(e_1) \subseteq R \wedge \forall e'[e_1 \sqsubset e' \rightarrow \neg(\{w_1\} \subseteq p)]$
$\vee\{w_3\} \subseteq p \wedge \tau(e_2) \subseteq R \wedge \forall e'[e_2 \sqsubset e' \rightarrow \neg(\{w_3\} \subseteq p)]) \wedge \ldots$
$\equiv \neg\exists e[dox(a,e) \subseteq p \wedge \tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow \neg(dox(a,e') \subseteq p)]]$
$\wedge\forall e[\tau(e) \subseteq R \wedge \forall e'[e \sqsubset e' \rightarrow dox(a,e') \subseteq \neg p)] \rightarrow dox(a,e) \subseteq \neg p]$

Syntactic Freezing as Semantic Matching: Licensing Syntactic Deviation in Echo Questions *

Jing Ji

McGill University, Montreal, Quebec, Canada
jing.ji@mail.mcgill.ca

Abstract

Echo questions (EQs) with an interrogative base have been regarded as syntactically more constrained than declarative-based EQs by involving a frozen CP structure (Sobin, 2010). This paper argues against syntactic freezing and claims that both declarative- and interrogative-based EQs are uniformly licensed by entailment defined over information content. The seemingly freezing effect of EQs arises despite entailment when relevant domain alternatives of the antecedent are factored into calculating the meaning of the base such that strict entailment is blocked by exhaustivity inference.

1 Introduction

This paper focuses on constituent echo questions (EQs) such as (1B) and (2B) with echo wh-words in capitals, which mark an information gap between the speaker and the hearer. The *antecedent* of an EQ is usually the immediately preceding utterance, which contains an element that can fill in the gap, deriving the *base* utterance of the EQ. Reis [2017] claims that EQs are not a clause type of their own but are dependent on their base structure types. Accordingly, (1B) is a *declarative-based EQ*, whereas (2B) is an *interrogative-based EQ*.

(1) A: Mozart was arrested by the police. B: The police arrested WHO?
base of (1B): The police arrested Mozart.

(2) A: Did Mary meet Putin? B: Did Mary meet WHO?
base of (2B): Did Many meet Putin?

(1) also shows that the base of an EQ can syntactically deviate from its antecedent. There has been debate about how EQs are related to their antecedents. Quotative analyses [Janda, 1985, Sobin, 2010] assume phonetic or syntactic copy involved in EQs. This view has been criticized by semantic analyses [Artstein, 2002, Poschmann, 2018, Beck and Reis, 2018], which maintain that EQs are related to their antecedents by way of content. One piece of evidence comes from the fact that some declarative-based EQs such as (3B) have no linguistic element in common with their antecedents. Interrogative-based EQs, however, are claimed to be subject to an additional condition, requiring antecedents with identical clause-type features. The minimal pair in (4) and (5) shows that a mismatch in clause type results in infelicity. This additional syntactic constraint for interrogative-based EQs is nevertheless unmotivated for semantic analyses and has been left as a puzzle.

(3) A: Did Lisa already say something about how her son likes MIT?
B: Tom is studying WHERE? (Beck and Reis, 2018)

*For helpful discussions, I thank Bernhard Schwarz, Michael Wagner, Syntax/Semantic reading group at McGill, the audience at the Toronto-Ottawa-Montreal Semantics Workshop, and anonymous abstract reviewers.

(4) A: Who will cook lentil souffle? B: Who will cook WHAT? (Beck and Reis, 2018)

(5) A: I wonder which unfortunate individual was told to prepare lentil souffle.
B: # Who will cook WHAT? (Beck and Reis, 2018)

Sobin [2010] explains the syntactically more constrained interrogative-based EQs and less constrained declarative-based EQs as a COMP-freezing effect, which requires a frozen CP structure (Spec CP, C, C') of the antecedent being copied. (6)-(8) illustrate COMP-freezing effect. (6A) is a declarative sentence, whose complementizer differs from that of interrogatives in (7A) and (8A), as manifested by subject auxiliary inversion. (8A) further differs from (7A) in that its Spec CP is filled with a wh-phrase. When serving as antecedents, they only allow EQs with matched CP structures, such as (6B), (7B), and (8B). EQs with unmatched CP structures as in (6B'), (7B'), and (8B') are infelicitous.

(6) A: Mary had tea with Cleopatra. (Sobin, 2010)
B: Mary had tea with WHO?
B': # Did Mary have tea with WHO?

(7) A: Did Mary have tea with Cleopatra? (Sobin, 2010)
B: Did Mary have tea with WHO?
B': # Mary had tea with WHO?

(8) A: What did Dracula drink at Mary's party? (Sobin, 2010)
B: What did WHO drink at Mary's party?
B': # Who drank what at Mary's party?

By contrast, Sobin considers non-CP elements as involving a loose copy. A consequence is that active–passive alternation is possible for declarative-based EQs but not interrogative-based EQs as voice change in interrogatives will be reflected in the C domain through head movement.

This paper argues against COMP-freezing by providing a unified semantic matching condition for licensing syntactic deviation in EQs. We show that both declarative- and interrogative-based EQs are licensed by entailment defined over information content between the antecedent and the base utterance. The seemingly freezing effect arises despite entailment since strict entailment between questions is always blocked by exhaustivity inference when the relevant domain alternatives of the antecedent are factored into calculating the meaning of the base.

The rest of the paper is organized as follows. Section 2 discusses the problem of COMP-freezing. Section 3 claims that Poschmann's (2018) entailment condition uniformly licenses declarative- and interrogative-based EQs and solves the problematic cases for Sobin. Section 4 illustrates that entailment is insufficient and requires enrichment, and demonstrates how the seemingly syntactic freezing effect is explained by the enriched semantic condition.

2 Problems of COMP-freezing

As a quotative analysis, COMP-freezing has been challenged by previous literature in that there are well-formed EQs without a frozen CP structure. As in (9B), the echo wh-word occurring in the CP structure makes it different from the antecedent. (10) provides another counterexample. While the antecedent in (10A) is a yes/no question with a complementizer triggering T to C movement, the EQ in (10B) is declarative-based. The well-formedness of (9B) and (10B) regardless of having uncopied CP elements shows that COMP-freezing is too strong.

(9) A: Which books about linguistics sell like mad?

B: Which books about WHAT sell like mad? (Beck and Reis, 2018)

(10) A: Could you pass me the salt?
B: You are asking me to hand WHAT? (Poschmann, 2018)

Moreover, COMP-freezing predicts that active–passive alternation is not allowed in interrogative-based EQs, but this is not the case in (11).

(11) A: Who arrested the thief? B: Who was WHO arrested by?

On the other hand, COMP-freezing is too weak in that it does not capture the semantic relation between EQs and their antecedents. For example, (12B) is a declarative-based EQ that only involves loosely copied non-CP elements. However, When we switch the order of the antecedent and the base utterance, as in (13), the EQ becomes infelicitous, indicating that there are semantic relations at play.

(12) A: Mary speaks French. B: WHO speaks a foreign language?

(13) A: Mary speaks a foreign language. B: # WHO speaks French?

In sum, COMP-freezing is neither necessary nor sufficient in licensing well-formed deviations in EQs. We argue in Section 3 that the problems above can be addressed by entailment.

3 Antecedent through entailment

Based on Schwarzschild's (1999) entailment licensing, Poschmann [2018] proposes that a declarative-based EQ is licensed only if the meaning of its base utterance is entailed by the information content of its antecedent, where the information content (IC) of an utterance φ in a given context c consists of the propositional content, its presuppositions PRES and its implicatures IMP [Van der Sandt, 1991], represented as $\mathrm{IC}(\varphi, c) = \{[\![\varphi]\!]^c \cup \mathrm{PRES}(\varphi, c) \cup \mathrm{IMP}(\varphi, c)\}$.

(14)-(16) demonstrate that declarative-based EQs are licensed by entailment.The base utterances of the EQs in (14B), (15B), and (16B) are entailed by the proposition content of (14A), the presupposition of (15A), and the implicature of (16A), respectively[1].

(14) A: Mary speaks French. B: WHO speaks a foreign language?
$[\![(14A)]\!]^c \Rightarrow [\![\textit{Mary speaks a foreign language}]\!]^{c'}$

(15) A: Bill knows that Mary speaks French. B: WHO speaks a foreign language?
$\mathrm{PRES}((15A), c) \Rightarrow [\![\textit{Mary speaks a foreign language}]\!]^{c'}$

(16) A: Jill may play the piano or the violin. B: WHO may play a string instrument?
$\mathrm{IMP}((16A), c) \Rightarrow [\![\textit{Jill may play a string instrument}]\!]^{c'}$

Different from Beck and Reis's (2018) hypothesis that interrogative-based EQs require additional syntactic constraint, we claim that Poschmann's (2018) entailment condition also holds for interrogative-based questions such that both declarative- and interrogative-based EQ are licensed in the same way. Guerzoni and Sharvit [2007] discusses entailment between questions based on speakers' intuition such that a question A entails a question B only if asking A automatically leads to asking B. Accordingly, there is no entailment in the question nucleus. As in (17), the original entailment relation ($[\![\textit{bought a red car}]\!] \Rightarrow [\![\textit{bought a car}]\!]$) disappears in the question nucleus since the speaker asking *Which student bought a car* can have no interest in the

[1] We assume that antecedents are uttered in context c while EQs are uttered in context c'.

color of the car, and vice versa. In contrast, there is downward entailment in the wh-restrictor of a wh-question. As in (18), the direction of entailment ($[\![\textit{French students}]\!] \Rightarrow [\![\textit{students}]\!]$) is reversed in the questions as asking *Which students passed the exam* automatically leads to asking *Which French students passed the exam.*

(17) No entailment in the question nucleus:
$[\![\textit{Which students bought a car}]\!] \not\Leftrightarrow [\![\textit{Which students bought a red car}]\!]$

(18) Downward-entailment in the wh-restrictor:
$[\![\textit{Which students passed the exam}]\!] \Rightarrow [\![\textit{Which French students passed the exam}]\!]$

Assuming entailment relation between questions as above, we demonstrate that the entailment condition defined over information content also licenses interrogative-based EQs. As in (19), the base of (19B) is asymmetrically entailed by its antecedent in (19A). In (20), the EQ is licensed by having a base entailed by the implicature of the antecedent. The fact that (19B) and (20B) are felicitous without strictly copied CP structure suggests that it is semantic relation rather than COMP-freezing that licenses syntactic deviation in EQs.

(19) A: Which French and German students passed the exam?
B: Which French students passed WHAT?

(20) A: I am curious about whether John speaks Wolof. B: Does John speak WHAT?

With entailment condition, we can understand why the prediction of COMP-freezing on voice alternation holds in some cases but not others. For instance, the infelicitous voice alternation in (8), repeated in (21), is attributed to no entailment relation between (21B') and IC of (21A). Conversely, when entailment is established between the base and antecedent, as in (11), repeated in (22), active–passive alternation is allowed.

(21) A: What did Dracula drink at Mary's party? (Sobin, 2010)
B': # Who drank what at Mary's party?

(22) A: Who arrested the thief? B: Who was WHO arrested by?

Also, those counterexamples to COMP-freezing pointed out by Beck and Reis [2018] and Poschmann [2018], repeated below, are no longer problematic with entailment condition. (23B) is licensed because its base utterance is semantically equivalent to the antecedent. The mismatched CP structure is allowed in (24) because the antecedent implicates a proposition that entails the base of (24B) such that our semantic condition is satisfied.

(23) A: Which books about linguistics sell like mad?
B: Which books about WHAT sell like mad? (Beck and Reis, 2018)

(24) A: Could you pass me the salt?
B: You are asking me to hand WHAT? (Poschmann, 2018)

4 Entailment is necessary but insufficient

Though the entailment condition is better than COMP-freezing in predicting possible syntactic deviation in EQs, it does not fully explain how this COMP-freezing illusion occurs as entailment itself is independent of clause types, shown in (25). This section discusses cases where the seemingly freezing effect occurs regardless of entailment. However, those cases are not counterexamples of entailment but suggest factors other than entailment at play. The enriched

semantic condition with those factors not only explains how the freezing effect arises but further demonstrates that declarative- and interrogative-based EQs are uniformly licensed.

(25) (Suppose Mary is a student.)
$[\![$*Which students passed the exam*$]\!] \Rightarrow [\![$*Did Mary pass the exam*$]\!]$

Despite the well-formed example in (19), we observe that asymmetric entailment between questions is always blocked in EQs. There are two kinds of strict entailment allowed between questions. One is downward-entailment in the wh-restrictor of wh-questions, as discussed in Section 3. The other is entailment from a wh-question to a polar question, as in (25). However, neither of them is sufficient to license EQs, as in (26) and (27). Not only questions, we observe that downward-entailment is not sufficient to license declarative-based EQs as well, as in (28).

(26) A: Which students passed the exam? B: # Which French students passed WHAT?

(27) A: Which students passed the exam? B: # Did Mary pass WHAT?

(28) A: Mary doesn't speak any foreign language. B: # WHO doesn't speak Igbo?

Before addressing the examples above, I will show that exhaustification of the base utterance triggered by alternatives plays a role in licensing EQs. Two relevant examples are shown below. The bases of (29B) and (30B) are strictly entailed by their antecedents, but both are infelicitous due to incompatible exhaustivity inference triggered by lexically determined scalar (σ-)alternatives associated with the base utterances.

(29) A: Jo teaches French and English. B: # WHO teaches French or English?

(30) A: John has three kids. B: # WHO has two kids?

To illustrate, consider (29B). The disjunctive connective *or* of the base utterance activates σ-alternatives as in (31). Those alternatives grow point-wise and are obligatorily factored into meaning via a covert exhaustifier O (modeled on *only*), deriving exhaustivity inference by negating all the contextually relevant alternatives (represented as C) that are excludable. In this case, we have a strengthened meaning of the base utterance as in (32), which implies that Jo does not teach both French and English. But this implicature is incompatible with what the antecedent asserted, and thus leads to infelicity.

(31) σ-ALT(or) = $\{\lambda x \lambda y.x \vee y,\ \lambda x \lambda y.x \wedge y\}$

(32) O_C(Jo teaches French or English) = teach(J,F)$\vee$teach(J,E)$\wedge\neg$(teach(J,F)$\wedge$teach(J,E))

We propose that the infelicity of (26)-(28) is also attributed to incompatible implicatures of the base, but rather than σ-alternatives, the domain for exhaustification contains domain (D-)alternatives activated by the antecedent. Chierchia [2013] assumes that a quantificational determiner carries a domain variable D, interpreted via an assignment function g and that the D-alternatives of the quantifier are derived by assigning a value that is a subset of $g(D)$ to this variable. Take (28) as an illustration. Assume D-alternatives activated by *any* in (33), with $g(D)$ consisting of three languages, Igbo (i), Spanish(s), and Arabic(a). With point-wise composition, we have a set of D-alternatives of the antecedent as in (34). Assuming those alternatives are inherited by the EQ and factored into calculating the meaning of the base, the exhaustivity inference is derived by negating all the relevant alternatives that are not entailed by the base. As in (35), the enriched meaning of the base implies that Mary speaks foreign languages other than Igbo, which contradicts the meaning of the antecedent.

(33) D-ALT(any$_D$) = $\{\lambda P \lambda Q.\exists x \in D'[P(x) \wedge Q(x)] : D' \subseteq \{i, s, a\}\}$

(34) D-ALT(Mary doesn't speak any$_D$ foreign language) = {¬spk(M,i∨s∨a), ¬spk(M,i∨s), ¬spk(M,i∨a), ¬spk(M,s∨a), ¬spk(M,i), ¬spk(M,s), ¬spk(M,a)}

(35) O$_C$(Mary doesn't speak Igbo) = ¬spk(M,i) ∧ spk(M,a) ∧ spk(M,s)

For questions that don't have a truth value such as (26), we assume that those D-alternatives project to a higher layer (over speech acts) such that exhaustivity inference operates over a set of question acts. Consider (26). Suppose there are two French students Mary and Fred, and a German student Sally such that D-alternative activated by *which* is in (36). Those alternatives grow point-wise into a set of question acts, as in (37), which serve as the relevant domain for exhaustification of the base. Thus the meaning of base is strengthened by implying that the speaker does not ask whether Sally passed the exam, which contradicts the antecedent. Since strict entailment for interrogative-based EQs always involves a base that is a domain alternative to the antecedent, the seeming freezing effect arises as a result of entailment failure.

(36) D-ALT(which$_D$) = $\{\lambda P \lambda Q.\exists x \in D'[P(x) \wedge Q(x)] : D' \subseteq \{Mary, Fred, Sally\}\}$

(37) D-ALT(which$_D$ students passed the exam) = $\{\text{QUEST}(\lambda p \exists x \in D'[\text{student}(x) \wedge p = \text{pass}(x, \text{exam})]) : D' \subseteq \{Mary, Fred, Sally\}\}$

(38) O$_C$(which French students passed the exam) =
QUEST(⟦*which French students passed the exam*⟧) ∧ ¬QUEST(⟦*Did Sally pass the exam*⟧)

In contrast, strict entailment is not blocked in (19), repeated below. Following Heycock and Zamparelli [2005], the domain of quantification of the question in (39A) is composed of a set of pluralities that each contain both French and German students. Hence the quantification domain of the base in (39B) is not a domain alternative to that of the antecedent such that no exhaustivity inference is derived. An alternative explanation is to regard (39A) as a conjunction of two questions by ellipsis such that (39B) has an antecedent semantically equivalence to the base. Evidence can be seen from the fact that EQs echoing one of the argument conjuncts of the antecedent are only acceptable when the predicate is distributive, e.g., (39) vs. (40).

(39) A: Which French and German students passed the exam?
B: Which French students passed WHAT?

(40) A: Which French and German students formed a circle?
B: # Which French students formed WHAT?

The contrast between (26) and (39) can also be accounted for by presupposition failure. The base of (26B) presupposes that some French students passed the exam. This presupposition is neither presupposed by the antecedent nor given in the preceding context. As EQs ask for closed information, we assume that ungiven presupposition cannot be accommodated and thus leads to infelicity. Similarly, presupposition failure also results in the infelicity of declarative-based EQs, as in (41), where the base carries an ungiven existential presupposition. (41) further confirms that declarative- and interrogative-based EQs are subject to similar constraints.

(41) A: Bill didn't get the perfect score. B: ?? It wasn't WHO who got the perfect score?

Given the above, the unified semantic licensing condition is enriched as follows.

(42) An utterance ω in context c' is a felicitous EQ to an utterance ϕ in context c iff its base (b(ω)) is entailed by IC(ϕ,c), and PRES(b(ω),c') and IMP(b(ω),c') are given in c.

References

Ron Artstein. A focus semantics for echo questions. In *Workshop on information structure in context*, volume 98, page 107. IMS, University of Stuttgart, 2002.

Sigrid Beck and Marga Reis. On the form and interpretation of echo wh-questions. *Journal of Semantics*, 35(3):369–408, 2018.

Gennaro Chierchia. *Logic in grammar: Polarity, free choice, and intervention*. OUP Oxford, 2013.

Elena Guerzoni and Yael Sharvit. A question of strength: on npis in interrogative clauses. *Linguistics and Philosophy*, 30(3):361–391, 2007.

Caroline Heycock and Roberto Zamparelli. Friends and colleagues: Plurality, coordination, and the structure of dp. *Natural language semantics*, 13(3):201–270, 2005.

Richard D Janda. Echo-questions are evidence for what? in papers from the general session at the twenty-first regional meeting. In *CLS. Papers from the General Session at the... Regional Meeting*, volume 21, pages 171–188, 1985.

Claudia Poschmann. Focus on repetition: on the role of focus and repetitions in echo questions. *Exact Repetition in Grammar and Discourse*, 323:295, 2018.

Marga Reis. On the analysis of echo questions. *Florida Linguistics Papers*, 4(3), 2017.

Roger Schwarzschild. Givenness, avoidf and other constraints on the placement of accent. *Natural language semantics*, 7(2):141–177, 1999.

Nicholas Sobin. Echo questions in the minimalist program. *Linguistic Inquiry*, 41(1):131–148, 2010.

Rob Van der Sandt. Denial. In *Chicago Linguistic Society*, volume 27, pages 331–344, 1991.

Plural priming revisited: *inverse preference* and *spillover effects**

Yizhen Jiang[1], Rebecca S. Ren[1], Yihang Shen[1], Richard Breheny[1], Paul Marty[2], and Yasutada Sudo[1]

[1] University College London, London, UK
[2] University of Malta, Msida, Malta

Abstract

Maldonado et al. (2017) observe that the distributive interpretation of sentences involving multiple plural expressions gives rise to stronger priming effects than their cumulative interpretation, and propose to interpret this observation in terms of structural priming of the phonologically silent distributivity operator. We report on a new experiment that included an additional 'neutral' baseline condition, whose results reveal that (i) the observed priming effects are inverse preference effects in that only the less dominant reading in the baseline condition gives rise to sizable priming effects, and (ii) both distributive and cumulative interpretations can have priming effects, depending on speakers' baseline preferences. We argue that these findings undermine Maldonado et al.'s claim that their results evidence the existence of the silent distributivity operator in syntax.

1 Introduction

Sentences that involve multiple numerical expressions like (1) can receive (at least) two readings, a cumulative reading and a distributive one, as exemplified below.

(1) Two boys have three balloons.
   a. **Cumulative reading**: There are two boys who, between them, have three balloons. Each boy has at least one balloon, and each balloon is owned by at least one boy.
   b. **Distributive reading**: Two boys have three balloons each.

The derivation of a (phrasal) distributive reading like (1-b) is standardly assumed to involve a silent *distributivity operator*. For cumulative readings, on the other hand, a number of different compositional semantic theories have been put forward (e.g., [3, 11, 15]).

Psycholinguistic studies have raised evidence that distributive readings are generally less preferred and involve more processing effort in comparison to cumulative and other non-distributive readings [1, 4, 5, 10, 7, 16, 17]. Particularly relevant for the present study is [12], which used priming. Their experimental results show (a) that sentences that are ambiguous between distributive and cumulative readings are more likely to be interpreted on the distributive reading after priming trials that force it (distributive primes) in comparison to the unprimed target trials that involve no priming (68.33% vs. 61.15%); and (b) that after priming trials that force the cumulative reading (cumulative primes), the rate at which the distributive reading is observed is not significantly different from the unprimed trials (58.17% vs. 61.15%). Based on these results, they make two claims: (i) The observed priming effects cannot be characterized

*We are grateful to Mora Maldonado for sharing with us the materials used in Maldonado et al.'s (2017) experiments. Data and analysis code are available at https://github.com/Yizhen-Jiang/plural_priming_revisited. This research was supported by the Leverhulme Trust Grant RPG-2018-425.

as inverse preference effects since the distributive reading was relatively preferred in the unprimed condition (61.15%); and (ii) the observed priming effects are to be explained in terms of structural priming of the silent distributivity operator, and as such constitute evidence for its existence.

[12]'s experimental design, however, has one potential issue:: They obtained their baseline results from trials that were interspersed between trials that involved priming, but as previously pointed out for implicature priming [13, 18], such baseline trials are potentially affected by preceding priming trials and do not directly inform us about neutral preferences.

In the present study, we report on an experiment that is just like [12]'s except that it has more neutral baseline conditions. Specifically, following [13, 18], we adopted a block design where the baseline trials with no priming were all conducted in the first block of the experiment, and priming was only introduced in in the second block. These novel baseline conditions enabled us to obtain data on participants' natural interpretative preferences. The results indicate that the observed priming effects are actually inverse preference effects, contrary to contradicting [12]'s claim (i) mentioned above. Furthermore, we observe that different speakers had different baseline preferences, and depending on the baseline preferences, both distributive and cumulative readings gave rise to priming effects. We argue that this finding undermines [12]'s claim (ii).

2 Experiment

Our experiment was built upon [12]'s materials and design while adding a more neutral baseline condition in order to investigate the direction of the priming effects observed in [12], which [12] could only speculate about in the absence of proper baseline. To achieve this, we adopted a two-block design: Block 1 contained no priming trials at all, allowing us to establish participants' interpretative preferences without potential spillover effects; Block 2 contained the same trials as in [12]'s Experiment 1. With this two-block design, the baseline condition can avoid contamination effects due to spillover from prime trials.

2.1 Participants

75 self-reported native speakers of English (36 females, 38 males, 1 other; average age 40.3 years old) were recruited via Prolific (UK/US IP addresses; prior approval rate $\geq 90\%$). Three participants were excluded prior to analyses: two for taking the survey on mobile phones rather than computers and one for suspiciously short mean response times (lower than 1s per trial). The survey took around 14 minutes to complete and participants were paid £2.4 for their time. All participants gave written informed consent.

2.2 Materials and Design

Our experimental materials were identical to Experiment 1 of [12]. Each trial consisted of one sentence and two pictures as exemplified in Figure 1. All sentences in the trials were generated based on one of the two frames as (2). Terms used for [shape 1] and [shape 2] featured the following shape names: heart(s), square(s), triangle(s), and circle(s). In each sentence, the name for [shape 1] was always different from that of [shape 2]. As reported in [12], the sentence list was generated in Python by randomly inserting shape words. To illustrate, Figure 1 presents schematized primes and targets used in the experiment.

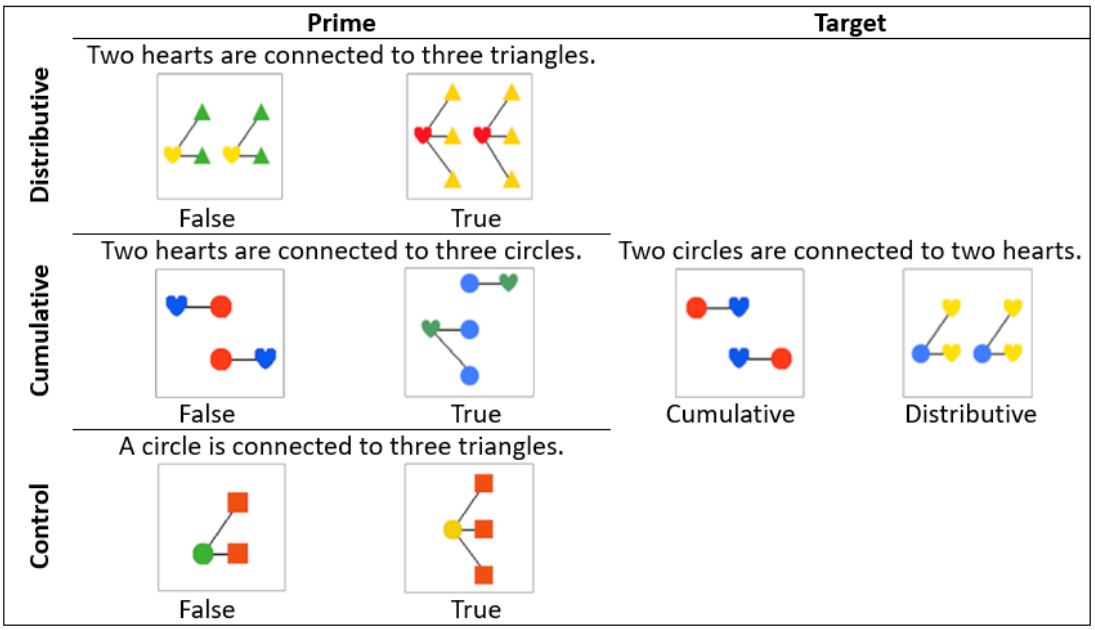


Figure 1: Example prime and target trials in the Cumulative, Distributive, and Control conditions in Block 2

(2) a. Two [shape 1] are connected to {two/three} [shape 2]. Ambiguous
b. A [shape 1] is connected to {two/three} [shape 2]. Unambiguous

Prime trials in block 2 consisted of an ambiguous sentence and two pictures: one picture was consistent with one of the two readings (cumulative or distributive) while the other was false because of the wrong number of shapes regarding of sentence. Control trials (called 'baseline' trials in [12]) consisted of an unambiguous sentence together with two pictures: one picture was making the sentence true while the other was making it false. As for target trials, sentences were all ambiguous. In the meantime, the two pictures were both "true" as they each corresponded to one of the readings of this sentence. Filler trials were similar to prime trials, except the false picture employed a different visual display.

Block 2 intended to replicate the original experiment in [12]. trials in this block followed a triplet pattern, meaning there were always two cumulative primes, two distributive primes, or two control primes before each target trail. There were 48 triplets, meaning 144 experimental trials in total. Another 48 filler trials were added, which made the total number of trials in this block achieve 192.

Trials in block 1 were extracted from trials in block 2. Target trials in this block were the same as target trials in block 2, which consisted of one ambiguous sentence and two pictures (one was consistent with cumulative reading, and another was consistent with distributive reading). At the same time, filler trials were the same as control trials of block 2.

Block 1 included 16 target trials, and 16 filler trials were inserted randomly. In this block, we intended to establish a neutral baseline without any possible priming effects. Participants could choose the picture according to their interpretations of sentences. In this block, different participants might exhibit different preferences between the cumulative reading and distributive

reading. Whether they were pro-cumulative or pro-distributive people, their statistics would serve as a better baseline for assessing priming effects' direction.

2.3 Procedure

Participants were instructed to click on the picture which they consider to be a better match for the sentence. Then followed a short practice phase where participants were given two practice trials, each of which contained an unambiguous sentence accompanied by two pictures. Feedback was given on participants' responses. Additionally, participants were asked to redo the trial if they select incorrect images. Participants could not enter the test phase until they gave correct responses on both practice trials. After the practice phase, the experiment began with the trials from Block 1 (Baseline conditions) and then continued with the trials in Block 2 (Distributive, Cumulative, and Control conditions).

2.4 Data treatment

Responses from one participant were removed from analyses because their mean accuracy rate of primes and fillers was below the pre-established 80% threshold. Following [12, 14], we further removed all responses to primed Target trials that were not preceded by the two correct prime responses. About 2% of all target responses were removed due to incorrect prime responses.

2.5 Results

The proportion of distributive choices on target trials by condition for all participants is displayed in Figure 2.5a. We conducted pairwise comparisons [9] between Distributive, Cumulative, and Control conditions in Block 2. Target responses from each group were analyzed by modelling response type likelihood using logit mixed effects models with the maximal random effects structure [2] which included random intercepts for subjects and items. The results of Block 2 showed essentially the same pattern as in [12]: Participants made significantly more distributive responses in the Distributive condition than in the Cumulative condition (62.0% vs. 54.8%; adjusted-$p < .001$) and no significant difference was detected between the Cumulative and Control conditions (54.8% vs. 59.1%; adjusted-$p = 0.154$). While the difference between the Distributive and Control conditions did not reach significance, unlike in Maldonado et al., the tendency in the same direction is present, and we consider this discrepancy between the two studies to be non-essential. This suggests participants' performance in the Priming block was largely unaffected by the prior Baseline block and hence secures our assessment of the participants' behavior in the Priming conditions.

Zooming in on the Baseline block, our analysis of the Baseline results using Hartigan's dip test of unimodality [8] identified two-mode distribution in participants' baseline preferences. This reveals that the majority of the participants had a preference for one of the two readings: Some participants (n = 22) tended to interpret the target sentences on their cumulative readings (henceforth, cumulative responders) while others (n = 45) favored the distributive interpretation (henceforth, distributive responders). The rest (n = 4) showed no preference.

In light of the inter-speaker baseline variation, we conducted further analysis of the results of Block 2 by sorting them according to the responder type (see Figure 2.5b). Pairwise comparisons [9] between Baseline, Distributive, Cumulative, and Control conditions were conducted for each responder group. Similar to the global analysis, target responses from each group were analyzed by modeling response type likelihood using logit mixed effects models [9] with random intercepts for subjects and items. The results reveal that the priming effects are **inverse**

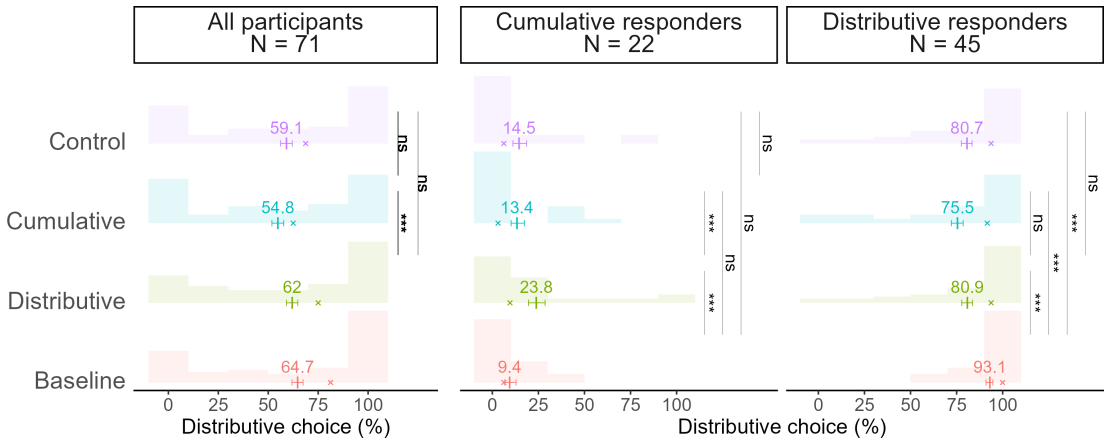


Figure 2: Proportion of distributive choices on target trials (a) by condition for all participants and (b) by responder group and condition. Histograms show the distributions of by-participant mean proportions. Crosses denote medians. Thick lines with numerals on top denote grand means. Error bars denote 95% binomial proportion confidence intervals. Asterisks denote significance levels based on the adjusted-p values for all comparisons.

preference effects in the following sense: For the cumulative responders, the priming effects were driven by the distributive primes boosting the rate of the distributive interpretation, while for the distributive responders, the priming effects were mainly driven by the cumulative primes inhibiting the rate of the distributive interpretation.

Thirdly, we observe spillover effects. Specifically, the distributive responders gave significantly fewer distributive responses in the Control condition than in the Baseline condition (80.7% vs. 93.1%; adjusted-$p < .001$), which confirms our suspicion that Maldonado et al.'s "baseline" results were not neutral, and affected by spillover effects from other trials in the same block. Furthermore, a similar difference was observed between the Distributive condition and the Baseline condition for the distributive responders (80.9% vs. 93.1%; adjusted-$p < .001$). This is consistent with the inverse preference pattern, where the distributive primes, forcing the favored interpretation for this type of participant, had no priming effect, and the difference from the Baseline condition was due to spillover effects. For the cumulative responders, we have a general tendency towards the opposite pattern, but the differences did not reach significance. We suspect this is partly due to the fact that there were relatively few cumulative responders in our data set.

Further, we detect no priming effects of the preferred interpretation on its own. Our results showed that the rate of distributive choice was about the same between the Control and Distributive condition for distributive responders and also between the Control and Cumulative condition for cumulative responders. It suggested that the effects observed in the Cumulative condition for cumulative responders were all attributed to the spillover effects from the distributive primes. Hence, cumulative primes could hardly modulate cumulative responders' interpretation. Similarly, distributive primes have no priming effects for distributive responders. That is, the favored interpretation in the Baseline condition did not have any observable priming effects by itself.

Finally, note that the size of the spillover effects is always smaller than the size of the priming effects. This is because spillover effects are in fact long-lasting remaining priming effects. Therefore, the difference between the Distributive and Cumulative conditions observed for all participants (interpreted as priming effects in [12] was the sum of the difference between the spillover effects and priming effects of Distributive primes for cumulative responders and that of Cumulative primes for distributive responders.

3 Discussion

Our results replicate the priming effects observed in [12] that participants provided more distributive responses after distributive primes than after cumulative primes. In addition, we found that these global effects are in fact caused by different prime types, depending on participants' baseline interpretive preferences. Specifically, distributive primes boosted the rate of distributive responses for cumulative responders while cumulative primes boosted the cumulative interpretation for distributive responders. We further observed that robust priming effects spilled over the targets in the other conditions within the same block: In both Control and Distributive conditions, cumulative responses were boosted by the lasting priming effects from cumulative primes in the Cumulative condition. Finally, we did not detect any priming effects of the less preferred interpretation: no additional effects were found in the Distributive condition relative to the Cumulative condition for distributive responders.

We argue that the above observations undermine [12]'s claim about the distributivity operator. Firstly, since the results exhibit an inverse preference pattern, we have to conclude that both distributive and cumulative primes have priming effects. If this observation were to be interpreted in terms of structural priming, we would conclude that both readings are to be accounted for by a silent operator (or at least some distinguished structural property such as silent movement), not just the distributive reading. Secondly, the inverse preference pattern is amenable to an alternative explanation without recourse to structural priming. Specifically, it could be accounted for in terms of online adaptation of probabilistic expectations about the distributions of the two readings, as proposed for certain other priming phenomena [6, 13]. Given this alternative explanation, we conclude that, although compatible with the existence of a silent distributivity operator, the results do not provide support for such an operator.

Our study replicates the results reported in [12]. In addition, our more neutral Baseline condition uncovered inter-participant variation in the default interpretation, which in turn revealed the inverse preference pattern in the priming results as well as spillover effects. This means that both distributive and cumulative interpretations can have priming effects, when they are the dispreferred reading. We pointed out that the observed priming effects as inverse preference effects can be explained without referring to structural priming, and therefore do not provide support for the presence of silent operators at LF.

References

[1] Stavroula Alexandropoulou, Jakub Dotlačil, Yaron Mcnabb, and Rick Nouwen. Pragmatic inferences with numeral modifiers: Novel experimental data. In *Proceedings of SALT 25*, pages 533–549, 2015.

[2] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278, 2013.

[3] Sigrid Beck and Uli Sauerland. Cumulation is needed: A reply to Winter (2000). *Natural Language Semantics*, 8(4):349–371, 2000.

[4] Patricia J. Brooks and Martin D. S. Braine. What do children know about the universal quantifiers all and each. *Cognition*, 60:235–268, 1996.

[5] Jakub Dotlacil. *Anaphora and distributivity: A study of same, different, reciprocals and others.* Netherlands Graduate School of Linguistics, 2010.

[6] Alex Fine, Ting Qian, T. Florian Jaeger, and Robert Jacobs. Syntactic adaptation in language comprehension. In *Proceedings of the 2010 Workshop on Cognitive Modeling and Computational Linguistics*, page 18–26, Uppsala, Sweden, Jul 2010. Association for Computational Linguistics.

[7] Lyn Frazier, Jeremy M Pacht, and Keith Rayner. Taking on semantic commitments, II: collective versus distributive readings. *Cognition*, 70:87–104, 1999.

[8] John A Hartigan and Pamela M Hartigan. The dip test of unimodality. *Annals of Statistics*, pages 70–84, 1985.

[9] T Florian Jaeger. Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4):434–446, 2008.

[10] Barbara Kaup, Stephanie Kelter, and Christopher Habel. Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes*, 17(4):405–450, 2002.

[11] Fred Landman. *Events and Plurality.* Springer, 2000.

[12] Mora Maldonado, Emmanuel Chemla, and Benjamin Spector. Priming plural ambiguities. *Journal of Memory and Language*, 95:89–101, 2017.

[13] Paul Marty, Jacopo Romoli, Yasutada Sudo, and Richard Breheny. Implicature priming, inverse preference, and context adaptation. Ms., UCL, 2022.

[14] Claudine N Raffray and Martin J Pickering. How do people construct logical form during language comprehension? *Psychological Science*, 21(8):1090–1097, 2010.

[15] Viola Schmitt. Pluralities across categories and plural projection. *Semantics and Pragmatics*, 12:17, 2019.

[16] Kristen Syrett and Julien Musolino Rutgers. Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, 20(4):259–291, 2013.

[17] Cherlon Ussery. Processing plural DPs: Collective, cumulative, and distributive interpretations. In *UMOP*, volume 37. GLSA, 2007.

[18] Brandon Waldon and Judith Degen. Symmetric alternatives and semantic uncertainty modulate scalar inference. In Stephanie Denison, Michael Mack, Yang Xu, and Blair C. Armstrong, editors, *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*, 2020.

The Acquaintance Inference as Scalar Exhaustification *

Jonathan Caleb Kendrick[1]

University of Maryland, College Park, Maryland, U.S.A.
jckend@umd.edu

Abstract

Predicates of personal taste (PPTs) trigger an inference that the speaker is acquainted with the thing being evaluated. While the acquaintance inference (AI) is often analyzed as part of an utterance's non-at-issue content, we argue it's best explained as a semantic entailment. Building on Magri's [15, 16] analysis of individual-level predicates as obligatory triggers of scalar exhaustification, we show that, if we treat the AI as arising from obligatory scalar exhaustification, we can explain much of the relevant projection data, provided we are willing to treating the AI as a semantic entailment of PPTs.

1 Introduction

Predicates of personal taste (PPTs) trigger an acquaintance inference (AI) that the speaker has first-hand knowledge of the object of predication [22, 3]. For example, *tasty* triggers an inference that the speaker has tried the thing in question.

(1) The cookies are tasty. $\rightsquigarrow$ *The speaker has tried the cookies.*

The AI has largely been viewed as part of an utterance's non-at-issue content [19, 17, 18, 2, 4, 12]. This paper argues that, in light of various projection data, we should take seriously the view, seldom defended in the literature, that the AI is a semantic entailment.[1]

We will assume that when a PPT appears without an overt experiencer, the PPT has a covert pronoun PRO in internal argument position.[2] Following Stephenson [21], we will assume that PRO refers to a judge parameter j. In autocentric contexts, the judge is the speaker. In non-autocentric contexts, it's some contextually salient individual. The view we'll defend in this paper is that a PPT relativized to PRO_j semantically entails that j has experienced the object in question.

(2) This cake is [tasty PRO_j] $\models$ PRO_j tried this cake

When a PPT occurs in the scope of conjunction, disjunction, and positive quantifiers, the AI behaves more-or-less like an ordinary entailment:

(3) a. The cookies and cake are tasty. $\rightsquigarrow$ *The speaker has tried the cookies and cake.*

b. The cookies or cake is tasty. $\rightsquigarrow$ *The speaker has tried the cookies or cake.*

*I would like to thank Fabrizio Cariani, Jéssica Mendes, Paolo Santorio, and Alexander Williams for their extensive feedback on earlier drafts of this squib. I would also like to thank the three anonymous referees for the Amsterdam Colloquium for their very useful comments which greatly improved the final version of this paper.

[1] A notable exception is Dinges and Zakkou [9] who argue that PPTs explicitly relativized to an experiencer (e.g., *the cookies are tasty to me*) entail their corresponding AI inferences.

[2] This is, I take it, is one of the dominant views in the literature [13, 21].

c. Every cookie is tasty. $\rightsquigarrow$ *The speaker has tried every cookie.*

However, unlike ordinary entailments, the AI persists in the scope of negation and negative quantifiers.[3]

(5) a. The cookies are not tasty. $\rightsquigarrow$ *The speaker has tried the cookies.*
b. None of the cookies are tasty. $\rightsquigarrow$ *The speaker has tried the cookies.*

In this respect, the AI patterns more like a presupposition, but, unlike standard presuppositions, which project from intensional environments, the AI is obviated by modals, attitude verbs, futurates, and questions.

(6) a. The cookies must be tasty. $\not\rightsquigarrow$ *The speaker has tried the cookies.*
b. I think the cookies are tasty. $\not\rightsquigarrow$ *The speaker has tried the cookies.*
c. The cookies are gonna be tasty. $\not\rightsquigarrow$ *The speaker has tried the cookies.*
d. Are these cookies tasty? $\not\rightsquigarrow$ *The speaker has tried the cookies.*

Likewise, the AI doesn't appear to be a conversational implicature, because it can't be canceled.

(7) ?? This cake is tasty, but I haven't tried.

So, if we can solve its negation problems, the entailment account seems like our best bet. In the next section, we sketch a theory on which PPTs trigger obligatory scalar exhaustification. This, we show, solves our initial negation problem, and yields many other plausible predictions.

2 PPTs as triggers of scalar exhaustification

Our positive theory starts from the common observation that PPTs behave like individual-level predicates (ILPs) [1, 19, 20]. Roughly, ILPs are predicates that pick out non-accidental properties (*tall*, *French*), unlike stage-level predicates (SLPs) which pick out accidental properties (*happy*, *sick*) [5]. ILPs have many well-known grammatical restrictions; for example, they do not tolerate temporal or locative modifiers [14].

(8) a. ?? John is tall in the classroom.
b. ?? John is French in the morning.
c. John is happy in the classroom.
d. John is sick in the morning.

[3] Dinges and Zakkou [9], who defend an entailment view like ours, attempt to explain this by positing that negation takes narrow scope with respect to a generic operator. The problem with this explanation is that we would then expect negation to sometimes take wide-scope over the generic. Dinges and Zakkou argue that we do see this:

(4) A: The cake tastes good to Hannah.—B: The cake doesn't taste good to Hannah. She hasn't even tried it!

However, we disagree that there's a wide-scope reading available here. If there is a felicitous reading of (4), it seems to be one where negation is interpreted meta-linguistically.

Magri [15, 16] explains the infelicity of (8a) and (8b) by positing that ILPs trigger obligatory scalar exhaustification, which is blind to conflicting information that's common ground. So, (8a) is obligatorily strengthened to 'John is only tall in the classroom' and (8b) is obligatorily strengthened to 'John is only French in the morning,' both of which are clearly infelicitous.

We propose that, as ILPs, PPTs also obligatorily trigger scalar exhaustification of a set of asymmetric alternatives. Following the "grammatical" approach to scalar implicature [6, 8], we'll treat LFs as containing a covert **exhaustification operator** (EXH) defined as follows:

$$[\![\text{EXH}]\!]^w = \lambda p\,.\, p(w) = 1 \;\&\; \forall q \in \mathit{Alt}(p)[q \in \mathcal{E}(p) \Rightarrow q(w) = 0]$$

where $\mathit{Alt}(\phi)$ is a set of alternatives to ϕ and $\mathcal{E}(\phi)$ is a subset of "innocently excludable" alternatives. The elements of $\mathit{Alt}(\phi)$ are the sentences ϕ' derivable from ϕ by successive replacement of scalar items with members of their Horn scale. The set of innocently excludable alternatives $\mathcal{E}(\phi)$ consists of all the $\psi \in \mathit{Alt}(\phi)$ such that:

- it's not the case that $\phi \wedge \neg\psi \models \bot$
- and there exists no $\chi \in \mathit{Alt}(\phi)$ such that $\phi \wedge \neg\psi \models \chi$ while $\phi \not\models \chi$.

In other words, the innocently excludable alternatives of ϕ are those alternatives whose negation is consistent with ϕ, and which don't allow one to infer alternatives not already inferable from ϕ alone.

When unnegated, the effect of exhaustifying *tasty* is vacuous and the AI is predicted to follow as an entailment; but, when negated, exhaustifying *tasty* leads to the scalar inference that the object in question is not tasty and that the speaker has tried it.

(9) a. $\mathcal{E}(\mathbf{tasty}(c)) = \emptyset$

b. $\text{EXH}([\![\text{the cake is tasty}]\!]) = \mathbf{tasty}(c)$

(10) a. $\mathcal{E}(\neg\mathbf{tasty}(c)) = \{\neg\mathbf{tried}(s, c)\}$

b. $\text{EXH}([\![\text{the cake is not tasty}]\!]) = \neg\mathbf{tasty}(c) \wedge \neg\neg\mathbf{tried}(s, c) = \neg\mathbf{tasty}(c) \wedge \mathbf{tried}(s, c)$

We've, therefore, solved our initial negation problem. In the following section, we show that this solution also makes correct predictions about extensional quantifiers, disjunction, and conjunction.

2.1 Extensional quantifiers

When a PPT occurs in the scope of a positive quantifier, the effect of exhaustification is vacuous and the AI projects as an ordinary semantic entailment.

(11) a. Some dessert is tasty $\rightsquigarrow$ *The speaker has tried some dessert.*

b. $\mathit{Alt} = \left\{ \begin{array}{l} \exists x[\mathbf{dessert}(x) \wedge \mathbf{tasty}(x)], \exists x[\mathbf{dessert}(x) \wedge \mathbf{tried}(s, x)] \\ \forall x[\mathbf{dessert}(x) \rightarrow \mathbf{tasty}(x)], \forall x[\mathbf{dessert}(x) \rightarrow \mathbf{tried}(s, x)] \end{array} \right\}$

c. $\text{EXH}([\![\text{some dessert is tasty}]\!]) = \exists x[\mathbf{dessert}(x) \wedge \mathbf{tasty}(x)]$

But, when in the scope of a negative quantifier, scalar exhaustification is non-vacuous, allowing us to correctly derive the AI as a scalar inference.

(12) a. No dessert was tasty. $\rightsquigarrow$ *The speaker has tried all the desserts.*

b. $Alt = \left\{ \begin{array}{l} \neg\exists x[\textbf{dessert}(x) \wedge \textbf{tasty}(x)], \neg\exists x[\textbf{dessert}(x) \wedge \textbf{tried}(s,x)] \\ \neg\forall x[\textbf{dessert}(x) \rightarrow \textbf{tasty}(x)], \neg\forall x[\textbf{dessert}(x) \rightarrow \textbf{tried}(s,x)] \end{array} \right\}$

c. EXH($[\![$no dessert was tasty$]\!]$) $= \neg\exists x[\textbf{dessert}(x) \wedge \textbf{tasty}(x)] \wedge \neg\neg\forall x[\textbf{dessert}(x) \rightarrow \textbf{tried}(s,x)] = \neg\exists x[\textbf{dessert}(x) \wedge \textbf{tasty}(x)] \wedge \forall x[\textbf{dessert}(x) \rightarrow \textbf{tried}(s,x)]$

So, we've also solved our negative quantifier problem.[4]

2.2 Disjunction and negated conjunctions

When a PPT occurs in a disjunction, the AI doesn't quite project like, say, a presupposition. However, when both disjuncts contain an occurrence of a PPT, the entire disjunction can license weak acquaintance-like inferences. For example, suppose we're at a party and you ask me which of the refreshments are good. I can't remember what I ate so I respond:

(14) Either the cake or the cookies are tasty. $\rightsquigarrow$ *The speaker tried the cookies or tried the cake.*

This is exactly what we would expect if the entailment account were correct. Each disjunct entails that the speaker has tried something–hence, the entire disjunction entails that the speaker has tried something. Our theory can also account for other nearby cases. For example, suppose we're at the party and I've tried something I didn't like. I can't remember what it was so I tell you:

(15) The cake and pie are not both tasty. $\rightsquigarrow$ *The speaker tried the cookies or tried the cake.*

This is something our theory captures as well:

(16) a. $Alt = \left\{ \begin{array}{l} \neg[\textbf{tasty}(p) \wedge \textbf{tasty}(c)], \neg[\textbf{tried}(s,p) \wedge \textbf{tried}(s,c)] \\ \neg[\textbf{tasty}(p) \vee \textbf{tasty}(c)], \neg[\textbf{tried}(s,p) \vee \textbf{tried}(s,c)] \end{array} \right\}$

b. EXH($\neg[\textbf{tasty}(p) \wedge \textbf{tasty}(c)]$) $= \neg[\textbf{tasty}(p) \wedge \textbf{tasty}(c)] \wedge \neg\neg[\textbf{tried}(s,p) \vee \textbf{tried}(s,c)] = \neg[\textbf{tasty}(p) \wedge \textbf{tasty}(c)] \wedge [\textbf{tried}(s,p) \vee \textbf{tried}(s,c)]$

The current theory can also explain why the AI disappears in excluded middle disjunctions like (18a).[5]

[4] Similarly, we also derive the correct prediction about negation in the scope of a universal quantifier:

(13) a. Every dessert is not tasty $\rightsquigarrow$ *The speaker has tried every dessert.*

b. $Alt = \left\{ \begin{array}{l} \exists x[\textbf{dessert}(x) \wedge \neg\textbf{tasty}(x)], \exists x[\textbf{dessert}(x) \wedge \neg\textbf{tried}(s,x)] \\ \forall x[\textbf{dessert}(x) \rightarrow \neg\textbf{tasty}(x)], \forall x[\textbf{dessert}(x) \rightarrow \neg\textbf{tried}(s,x)] \end{array} \right\}$

c. EXH($[\![$some dessert is tasty$]\!]$) $= \forall x[\textbf{dessert}(x) \rightarrow \neg\textbf{tasty}(x)] \wedge \neg\exists x[\textbf{dessert}(x) \wedge \neg\textbf{tried}(s,x)] = \forall x[\textbf{dessert}(x) \rightarrow \neg\textbf{tasty}(x)] \wedge \forall x[\neg\textbf{dessert}(x) \vee \textbf{tried}(s,x)]$

[5] Our theory also derives the correct predictions in nearby cases too like negated contradictions containing PPTs. For example, Cariani [4] observes that the AI is obviated in (17a).

(17) a. The cake is not both tasty and not tasty. $\not\rightsquigarrow$ *The speaker has tried the cake.*

(18) a. The cake is either tasty or not tasty. $\not\rightsquigarrow$ *The speaker has tried the cake.*

b. $Alt = \left\{ \begin{array}{l} \textbf{tasty}(c) \vee \neg\textbf{tasty}(c), \textbf{tried}(s,c) \vee \neg\textbf{tried}(s,c) \\ \textbf{tasty}(c) \wedge \neg\textbf{tasty}(c), \textbf{tried}(s,c) \wedge \neg\textbf{tried}(s,c) \end{array} \right\}$

c. $\textsc{exh}(\neg[\textbf{tasty}(c) \wedge \neg\textbf{tasty}(c)]) = [\textbf{tasty}(c) \vee \neg\textbf{tasty}(c)]$

In these contexts, the only innocently excludable alternative is a contradiction–e.g., $\textbf{tried}(c) \wedge \neg\textbf{tried}(s,c)$–so the effect of excluding it is vacuous.

While excluded middle contexts are perhaps an extraordinary case, far more mundane disjunctions obviate the AI. For example, Ninan imagines that we are in a restaurant, about to order, and everyone around us is eating lobster rolls [18, pg. 774]. In this context, the following utterance does not trigger the AI:

(19) Either the lobster rolls here are delicious or they're out of everything else.

Given our theory, we correctly predict that the AI does not project in (19). However, our theory still appears to make a problematic prediction: since *delicious* entails trying, our theory appears to predict that (19) entails that either the speaker has tried the lobster rolls or the restaurant is out of everything else. This inference is clearly not licensed in the given context. However, we believe an acquaintance inference *is* licensed in this context, but not one where the experiencer is the speaker. For example, if the taste predicate is explicitly relativized to the speaker, the utterance becomes markedly worse:

(20) ?? Either the lobster rolls here are delicious to me or they're out of everything else.

However, if the taste predicate is explicitly relativized to, say, the patrons of the restaurant, the disjunction is felicitous, and the predicted entailment seems far more natural.

(21) Either the lobster rolls here are delicious to them (the patrons) or they're out of everything else.

Disjunction also provides further support for our hypothesis that *tasty* entails *tried*. Hurford observed that a disjunction $X \wedge Y$ is infelicitous if X entails Y or Y entails X [11]. Intriguingly, (22a) seems strikingly similar to paradigm violations of Hurford's constraint (22b).

(22) a. ?? Either John tried the burger or he found it tasty.

b. ?? Either John is in Paris or he's in France.

Gazdar [10] also observed that Hurford's constraint admits exceptions like (23).

(23) Mary read some or all of the books.

We can find parallel exceptions with PPTs too. For example, when we negate both disjuncts, the disjunction is no longer infelicitous (24a). Chierchia, Fox, and Spector [7, 8] suggest that,

b. $Alt = \left\{ \begin{array}{l} \neg[\textbf{tasty}(c) \wedge \neg\textbf{tasty}(c)], \neg[\textbf{tried}(s,c) \wedge \neg\textbf{tried}(s,c)] \\ \neg[\textbf{tasty}(c) \vee \neg\textbf{tasty}(c)], \neg[\textbf{tried}(s,c) \vee \neg\textbf{tried}(s,c)] \end{array} \right\}$

c. $\textsc{exh}(\neg[\textbf{tasty}(c) \wedge \neg\textbf{tasty}(c)]) = \neg[\textbf{tasty}(c) \wedge \neg\textbf{tasty}(c)]$

in Gazdar's exceptions, scalar exhaustification of one of the disjuncts breaks the entailment between them; hence, there is no violation of Hurford's constraint. If (24a) has the LF in (24b), as our theory predicts, then this would explain why (24a), but not (22a), is felicitous.

(24) a. Either John didn't try the burger or he didn't find it tasty.

b. Either John didn't try the burger or EXH[he didn't find it tasty].

2.3 Intensional operators

While the entailment view coupled with scalar exhaustification has made light work of the projection problem in extensional contexts, we still face difficulties in intensional contexts. As we noted in §1, virtually all intensional contexts–modals, conditionals, and attitude reports–obviate the AI. But, given, for example, the plausible assumption that knowledge is veridical, the entailment view incorrectly predicts that (25) should be a contradiction.

(25) I know that the cake is tasty, even though I haven't tried it, because every desert at the French Laundry is amazing.

We think that intensional operators obviate the AI, because they shift the judge parameter.

$$[\![know]\!]^{w,t,j} = \lambda p.\forall \langle w', t', x' \rangle \in Epi_{w,t,j} : p(w')(t')(x') = 1$$

So, (25) isn't a contradiction, because the attitude verb *know* shifts the judge parameter to a non-speaker individual. This response seems plausible. After all, if we explicitly specify that the judge is the speaker, then we get something that seems contradictory:

(26) ?? I know that the cake is tasty to me, but I haven't tried it.

In short, we believe the AI isn't completely obviated by intensional contexts–rather, intensional contexts often generate the appearance of obviation by shifting the judge parameter from the speaker or attitude holder to another individual.

3 Conclusion

We've sketched how, by helping ourselves to a "grammatical" theory of scalar implicature, an entailment view of the AI might solve the projection problem for PPTs. In many ways, the AI behaves like an ordinary scalar inference, but with a few quirks. First, unlike ordinary conversational implicatures, the AI is obligatory and its computation is blind to information that's common ground. Secondly, by manipulating the judge argument of PPTs, intensional operators introduce noise into our data, obscuring the fact that the AI is simply a scalar inference.

References

[1] Pranav Anand. Kinds of taste. Manuscript, 2009.

[2] Pranav Anand and Natasha Korotkova. Acquaintance content and obviation. *Proceedings of Sinn und Bedeutung*, 22:55–72, 2018.

[3] Malcolm Budd. The acquaintance principle. *British Journal of Aesthetics*, 43(4):386–392, 2003.

[4] Fabrizio Cariani. *The Modal Future: A Theory of Future-Directed Thought and Talk*. Cambridge, UK: Cambridge University Press, 2021.

[5] Gregory Carlson. *Reference to Kinds in English*. PhD thesis, Amherst MA, 1977.

[6] Gennaro Chierchia. Broaden your views: implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, 37(4):535–590, 2006.

[7] Gennaro Chierchia, Danny Fox, and Benjamin Spector. Hurford's constraint and the theory of scalar implicatures. In Paul Egre and Giorgio Magri, editors, *Presuppositions and implicatures: Proceedings of the MIT–Paris Workshop*, pages 47–62. 2009.

[8] Gennaro Chierchia, Danny Fox, and Benjamin Spector. Scalar implicature as a grammatical phenomenon. In Klaus von Heusinger, Claudia Maienborn, and Paul Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, pages 3–2297. De Gruyter Mouton, 2012.

[9] Alexander Dinges and Julia Zakkou. Taste, traits, and tendencies. *Philosophical Studies*, 178(4):1183–1206, 2021.

[10] Gerald Gazdar. *Pragmatics: Implicature, Presupposition and Logical Form*. Academic Press, 1979.

[11] James R. Hurford. Exclusive or inclusive disjunction. *Foundations of Language*, 11(3):409–411, 1974.

[12] Christopher Kennedy and Malte Willer. Assertion, expression, experience. *Inquiry: An Interdisciplinary Journal of Philosophy*, 65(7):821–857, 2022.

[13] Ezra Keshet. A matter of taste. Manuscript, 2005.

[14] Angelika Kratzer. Stage level and individual level predicates. In Gregory Carlson and Francis Jeffry Pelletier, editors, *The Generic Book*, pages 125–175. University of Chicago Press, 1995.

[15] Giorgio Magri. A theory of individual-level predicates based on blind mandatory scalar implicatures. *Natural Language Semantics*, 17(3):245–297, 2009.

[16] Giorgio Magri. Another argument for embedded scalar implicatures basedon oddness in downward entailing environments. *Semantics and Pragmatics*, 4(6):1–51, 2011.

[17] Dilip Ninan. Taste predicates and the acquaintance inference. *Proceedings of Semantics and Linguistic Theory (SALT)*, 24(1):290–309, 2014.

[18] Dilip Ninan. The projection problem for predicates of taste. *Proceedings of Semantics and Linguistic Theory (SALT)*, 30(1):753–778, 2020.

[19] Hazel Pearson. A judge-free semantics for predicates of personal taste. *Journal of Semantics*, 30(1):103–154, 2013.

[20] Eric Snyder. Binding, genericity, and predicates of personal taste. *Inquiry: An Interdisciplinary Journal of Philosophy*, 56(2-3):278–306, 2013.

[21] Tamina Stephenson. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4):487–525, 2007.

[22] Richard Wollheim. *Art and its Objects*. Cambridge University Press, 1980.

Investigating the properties of clause-embedding predicates in Polish*

Tomasz Klochowicz

University of Amsterdam
klochowicztomasz@gmail.com

Abstract

This paper's main focus lies on the compositional semantics of clause-embedding predicates, i.e. verbs or verb-like expressions that represent a relationship between a subject and a proposition. It provides an improved characterisation of the selectional behaviour of the classes of responsive and anti-rogative predicates in terms of their semantic properties. I propose refinements of the hypotheses by Uegaki and Sudo [US19] and Roelofsen and Uegaki [RU21], that are falsified by empirical data. My new proposals are that *all non-veridical and positively preferential predicates are anti-rogative* and that *all responsive predicates are either Q-to-P or P-to-Q distributive*. The latter hypothesis can be alternatively formulated as: "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*".

1 Introduction

Our conversations frequently involve the use of clause-embedding predicates, i.e. verbs or verb-like expressions that represent a relation between a subject and a proposition. The compositional semantics of the clause-embedding has been intensively discussed in the recent literature in linguistics and formal semantics [e.g. SE15, TRA18, TRA$^+$19, Ueg19, RU21]. The main objective of studying these predicates is to identify the constraints that describe their selectional behaviour in terms of their semantic properties, i.e to explain why *responsive* predicates like "*to know*" can embed all types of clauses as in (1), but *rogative* predicates, like "*to wonder*" in (2) or *anti-rogative* like "*to believe*" in (3) can embed respectively only interrogative and only declarative clauses. Since clause-embedding predicates occur in many languages, the postulated constraints should be verified against cross-linguistic data.

(1) Ann **knows**...
   a. that Bob left.
   b. whether Bob left.
   c. who left.

(2) Ann **wonders**...
   a. *that Bob left.
   b. whether Bob left.
   c. who left.

(3) Ann **believes**...
   a. that Bob left.
   b. *whether Bob left.
   c. *who left.

As a contribution to this line of research, using the methodology by [URR$^+$22], I created a database of clause-embedding predicates in Polish containing more than 1500 data points and analysing 48 predicates with respect to 23 semantic and 13 selectional properties.[1] Moreover, I verified the constraints postulated in the literature against it. In the remainder of this paper, I will discuss the data collection method and the results of that procedure.

*I would like to thank my thesis supervisors: Floris Roelofsen and Jakub Szymanik for all the help and support in creating this project.

[1]The database is available at https://docs.google.com/spreadsheets/d/1qijQdak7ppgDqFB91iY-003P3MV-21Wq/edit?usp=sharing&ouid=117127311628580729011&rtpof=true&sd=true and is a part of my master thesis project [Klo22].

2 Some remarks on the data collection process

To collect the data I used the MECORE questionnaire by Uegaki et al. [URR+22]. It consists of a list of clause-embedding predicates in English and several linguistic tests which ought to be performed on each predicate to analyse its semantic and selectional properties. Moreover, the questionnaire provides a spreadsheet to summarise the collected data. Therefore data collection process can be described in three steps: 1. translation, 2. application of the tests, and 3. testing and adjusting the tests and predicates.

The database, which is the main output of this process is a table with properties as columns and predicates as rows, which allows to easily check the validity of a hypothesis describing the behaviour of the clause-embedding predicates. To evaluate a hypothesis in the form of an implication, we isolate all the verbs which satisfy the property from the antecedent and check whether they also satisfy the property from the consequent. To check the validity of a sentence in the form of a bi-implication, we do this twice, once for each implication. To assess the validity of a universal statement, we check whether all the verbs satisfy the relevant property. These procedures will automatically result in finding the relevant counterexamples if they exist. This method can be very fruitful and provide interesting results, which I discuss the next section. However, before proceeding to this part, I need to address its several shortcomings and issues.

The length of the list of predicates makes it infeasible to run a study with enough native speakers to ensure full-proof results. Therefore, the results of the study are questionable as they depend on the judgment of one person. Since the database contains around 1500 data points mistakes are statistically unavoidable. Moreover, the judgments frequently differ between native speakers, and this method will not detect any variation. To accommodate this issue, the primary investigators organised regular meetings where consultants working on data from different languages could compare their results. Obviously, cross-linguistic variations exist, but since the predicates have the same meaning, they should not be widespread. Therefore, looking at some language can provide insight into another, which helps avoid some mistakes.

The confirmation of the hypotheses, both existing in the literature and new ones is questionable. Since the database does not include all predicates and since only one consultant creates it, it is impossible to get an ultimate confirmation of a hypothesis even for one language. Moreover, because of these issues the actual degree of confirmation is rather low. However, the study can provide a cross-linguistic confirmation of the counterexamples of some hypotheses and strengthen their falsification or even allow to find new counterexamples as presented in the next section. Moreover, running a short study on a bigger sample can be quickly performed to confirm or reject the falsifying piece of evidence as it needs to consider only one predicate and a few relevant properties.

The generality of the claims in the questionnaire is questionable. By definition, any semantic property ranges over all possible subjects and all possible complements. The consultants of the questionnaire analyse only one specific example, which may lead to false conclusions for some verbs. Therefore a strong (but fairly standard) assumption is required, namely that the predicates behave uniformly concerning all the sufficiently similar subjects and complements [Klo22].

The scope of the study includes only some semantic and selectional properties. The choice of those properties is inspired by previous research. Therefore discovering a new hypothesis

in the process is unlikely, as the combinations of the properties have already been checked for English and other languages. Considering the most influential papers (e.g. [SE15], [TRA+19] or [RU21]), the most fruitful way to describe the behaviour of clause-embedding predicates is to look at the properties that haven't been considered before. Moreover, the study only analyzes the selectional properties of the verbs in simple contexts like $\ulcorner sVc \urcorner$ where s is a subject, V is a predicate and c is a complement of some type. Many verbs change their selectional properties in more complex contexts (under negation, modalities or tense/aspect change). For instance, the English predicate "*to doubt*" usually does not take interrogative complements, but in some negative contexts, it does as in (4). This behaviour should also be explained by the semantics theory of clause-embedding predicates.

(4) "Nobody doubts who the mayor of London is; nobody doubts who speaks for Scotland...".[2]

It should be clear by now that all the issues are interdependent. For instance, performing a study that would provide a higher degree of confirmation would drastically increase the size. Therefore the only reasonable approach is to keep all these three factors balanced, and such a balance seems to be achieved by the current method. Moreover, considering additional predicates, properties, or even contexts can be done without trouble, but it is impossible to include everything.

Is it, therefore, worth the effort to perform such a study? I would argue that it is. First of all, as mentioned above, the database is a good falsification tool. It allows detecting potential counterexamples and running a detailed study of these predicates to confirm the consultant's judgement. Thus the database should not be treated as an oracle that gives the researchers written-in-stone facts about language. However, it can provide a good starting point for further independent research. Therefore the size of the database can be seen as its feature because it allows looking at many predicates to determine which are worth to be investigated closely. The next section provides an example of the important conclusions, which can be drawn from the database.

3 Modelling selectional behaviour

In this section, I will discuss the results of the empirical study of clause-embedding predicates in Polish. I will present some insights into the veridicality hypothesis proposed by Paul Egré [Egr08], and afterwards, I will present empirical evidence which falsifies two other hypotheses previously postulated in the literature and propose their refinements.

3.1 The veridicality hypothesis

A predicate (V) is called *veridical* with respect to declaratives if and only if a sentence which relates a subject s to a declarative complement d using V implies the complement d. For instance, the English predicate "to know" is veridical with respect to declaratives as (5-a) implies (5-b):

(5) a. Ann knows that Peter left.

[2] https://www.theguardian.com/uk-news/2016/feb/11/labour-tony-lloyd-greater-manchester-mayoral-bid-devolution [accesed: 13.05.2022] Found using the *English corpus of News on the Web* [Dav16]. See also other examples for the predicates "*believe*", "*be certain*", "*think*", "*hope*" in the articles: [Rob19], [May19], [Özy21] and [Whi21].

b. Peter left.

The veridicality hypothesis $\mathbb{V}$ proposed by Paul Egré claims that all veridical predicates are responsive [Egr08]. Egré himself points out that emotive predicates like "*to regret*" are potential counterexamples to his hypothesis, as they seem to be veridical and not embed questions. Mayr adds the verb "*to resent*" to this list [May19].

However, Egré argues that these emotive predicates should not be seen as a counterexample to his hypothesis, as they only imply that the complement is believed by the subject and not that it is true [Egr08]. However, the evidence suggests that they pass the veridicality test so they can be seen as proper counterexamples. I will argue that further empirical research should be performed to check the validity of this hypothesis.

Intuitively veridical predicates should not allow for complement cancellation, as they imply that the complement is true. Therefore, the empirical study, except just checking the assertability of sentence (7-a), should also test whether sentences like (7-b) or (7-c) are well-formed. In these cases the intuitions of the native speakers are clear, and most of them claim that there is something wrong with these sentences, as they seem to be contradictory. However, some are still convinced that these sentences are more or less correct. Thus investigating these sentences may lead to interesting conclusions regarding the verb "*to regret*".

(6) Peter invited Anna on a date. His evil twin - Paul - decided to destroy his brother's relationship and showed up at the meeting with Anna instead of Peter. During the evening, Paul was very rude, and thus Ann regrets that this meeting happened. However, Ann is still convinced that she met with Peter, not Paul.

(7) a. Ann regrets going on a date with Peter.
b. ?Ann regrets going on a date with Peter, but in fact, she went on a date with Paul.
c. ??Ann went on a date with Paul, but she regrets going on a date with Peter.

Investigation of predicates like "*regret*" may also lead to interesting conclusions regarding the property of veridicality. As mentioned above, Egré's motivation to discuss this property is the fact that it is the main semantic difference between "*know*", which is responsive and "*believe*" or "*think*", which are anti-rogative in neutral contexts [Egr08]. It seems that "*regret*" lies somewhere on the spectrum, which starts with a non-veridical "*think*" and ends with the veridical "*know*". Perhaps it is at one extremum or the other, but the evidence suggests that it is somewhere between the two ends. The future study should consider it possible that "*regret*" may undermine the assumptions of the MECORE project, i.e. that all the semantic properties are binary.

3.2 Analysis and refinement of the $\mathbb{NVP}$ hypothesis

The $\mathbb{NVP}$ hypothesis proposed by Ueagki and Sudo claims that any predicate that is non-veridical and preferential is also anti-rogative [US19]. There are at least three predicates in Polish that are counterexamples to this hypothesis: "*martwić się*", "*bać się*" and "*narzekać*" (*eng*: "*be worried*", "*fear*" and "*complain*" respectively).

To derive the anti-rogativity of the non-veridical preferentials Ueagki and Sudo proposed that an utterance of such a predicate presupposes *Threshold Significance*, i.e. that "[there is] an element in the comparison class whose degree along the relevant scale [of preference] is higher than the threshold [of preference]"[US19, p.335]. Observe that, all the counterexamples violate this presupposition, as they are *negatively* preferential. For instance, in the context (8), the premise of the reasoning (9) is true, but the conclusion is not. Thus, since the predicate

to be worried can be meaningfully uttered in this context, it does not presuppose *Threshold Significance*.

(8) Peter is a prince who needs to marry the winner of the race. He knows nothing about the participants but it still matters to him who will win the race.

(9) Peter is worried who won the race.
$\not\rightsquigarrow$ Peter is worried that x_F won the race.

To accommodate for the issues with the $\mathbb{NVP}$ hypothesis, which are caused by predicates "martwić się", "bać się" and "narzekać", a different conclusion should be drawn from the derrivation by Uegaki and Sudo [US19]. Observe that the *Threshold Significance* presupposition coincides with the property of being *positively preferential*, which is that, in the declarative case, an agent prefers the clause embedded under the predicate over the other options. Therefore I argue that the condition of being *positive* should be added to the $\mathbb{NVP}$ hypothesis. The new claim is not violated by my counterexamples and ensures that the predicates that satisfy the antecedent of the hypothesis also satisfy the *Threshold Significance* presupposition. Thus I propose the hypothesis $\mathbb{NVPP}$ as a refinement of the hypothesis $\mathbb{NVP}$. $\mathbb{NVPP}$ is formally justified by the derivation in the aforementioned paper and independently confirmed by the data collected in my database:

Hypothesis 1. $\mathbb{NVPP}$*: All non-veridical and positively preferential predicates are anti-rogative.*[3]

3.3 The entailment hypothesis

The $\mathbb{PQ}$ hypothesis proposed by Roelofsen and Uegaki claims that all responsive predicates are P-to-Q distributive, i.e. that a sentence with a declarative complement embedded under a responsive predicate entails the same sentence but with a respective interrogative [RU21]. For example, the English predicate “to announce” is P-to-Q distributive as (10-a) implies (10-b). Q-to-P distributivity can be defined as an inverse of this implication.

(10) a. About some player a Alfred announced that a won the race.
b. Alfred announced who won the race.

The analysis of the database shows that there are verbs that are not P-to-Q distributive. Observe that the verb "*to learn*" seems to behave the same way in English. For instance, the sentence (11), which has a false declarative complement, can be meaningfully uttered, but (12) seems to presuppose that Ann learned the *true* answer to the question. Thus (11) is inconsistent with (12). Thus the predicate "*to learn*" violates the $\mathbb{PQ}$ hypothesis.

(11) At school, Ann learned that Benjamin Franklin was the first president of the USA.

(12) At school, Ann learned who was the first president of the USA.

The issue is even more visible in Polish. The context (13) clearly shows that a subject can *learn* false information and that even thought, the sentence (14) contains a question, which the sentence (13) answers, it is not assertible in this context.

[3] A similar hypothesis was also independently suggested in the GLOW presentation by Özyıldız et al. as a potential solution to the issue with negative preferential predicates [ÖQR+22].

(13) a. "Constantin Reliu dowiedział się w styczniu, że ... nie żyje. Próbuje co prawda udowodnić rumuńskim władzom, że jest inaczej, ale napotyka potężny opór biurokracji".[4]

b. In January, Constantin Reliu learned that ... he is dead. He tries to prove to the Romanian administration that it is not the case, but he is met with strong resistance on their side.

(14) Constantin dowiedział się czy nie żyje.
Constantin learned himself whether NEG he is alive.
"Constantin learned whether he is dead."

However, it is important to notice that in the database there is no predicate that would be neither Q-to-P nor P-to-Q distributive. Moreover, Spector and Egré independently claimed that the existence of predicates, that would have a completely unrelated meaning when they embed declaratives and interrogatives is implausible [SE15].[5] From this observation, they conclude that it is always possible to define the meaning of any sentence containing some embedded interrogative by the meaning of the respective declarative embeddings. I would argue that this observation cannot be strengthened like this and, based on the database, I would like to propose the following hypothesis:

Hypothesis 2. *(Entailment hypothesis) All responsive predicates are Q-to-P or P-to-Q distributive.*

This hypothesis can be alternatively formulated as: "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*" and as such it follows the idea of [SE15]. Moreover, is also a weakening of the $\mathbb{PQ}$ hypothesis and accommodates all the aforementioned counterexamples to it. In particular, it predicts the existence of predicates that are non-veridical but Q-to-P veridical, which are considered by Roelofsen and Uegaki to be potential counterexamples to their hypothesis (Buryat "*hanaxa*" and Turkish "*bil*") [RU21].[6]

4 Conclusion

In this paper, I presented a database of the clause-embedding predicates in Polish and the main conclusions arising from its analysis. I provided new linguistic evidence regarding the predicate "*to doubt*", discussed the non-standard behaviour of the predicate "*to regret*", and proposed two refinements of the hypotheses that were previously established in the literature. I argued that *all non-veridical and positively preferential predicates are anti-rogative* and that *all responsive predicates are either Q-to-P or P-to-Q distributive*, which can also be formulated as "sentences with corresponding interrogative and declarative complements embedded under the same predicate are always *related by entailment*." These hypotheses provide a better understanding of the behaviour of clause-embedding predicates and are a step towards their full unified theoretical analysis.

[4] https://www.rmf24.pl/fakty/swiat/news-w-styczniu-dowiedzial-sie-ze-nie-zyje-wladze-nie-chca-przyja,nId,2558031#crp_state=1 [access: 01.06.2022]

[5] e.g. their "*shknow*", which they define as "know" if it embeds a declarative and as "wonder" if an interrogative.

[6] However, as pointed out by [RU21] the predicate "*magtaka*" in the Tagalog language is translated to English as "to surprise" with declarative complements and as "wonder" while embedding interrogatives. A more detailed study of this predicate is needed to investigate whether it indeed violates $\mathbb{E}$.

References

[Dav16] Mark Davies. Corpus of news on the web (NOW). Available online at https://www.english-corpora.org/now, 2016.

[Egr08] Paul Egré. Question-embedding and factivity. *Grazer Philosophische Studien*, 77(1):85–125, 2008.

[Klo22] Tomasz Klochowicz. Investigating semantic and selectional properties of clause-embedding predicates in Polish. Available at: https://www.illc.uva.nl/Research/Publications/Reports/MoL/ in the *Master of Logic (MoL) Series*, 2022. Institute for Logic, Language and Computation, Universiteit van Amsterdam.

[May19] Clemens Mayr. Triviality and interrogative embedding: context sensitivity, factivity, and neg-raising. *Natural Language Semantics*, 27(3):227–278, 2019.

[ÖQR+22] Deniz Özyıldız, Ciyang Qing, Floris Roelofsen, Maribel Romero, and Wataru Uegaki. Cross-linguistic patterns in the selectional restrictions of preferential predicates. Presentation on a work in progress, 2022.

[Özy21] Deniz Özyıldız. *The event structure of attitudes*. PhD thesis, University of Massachusetts Amherst, 2021. Doctoral Dissertations. 2209.

[Rob19] Tom Roberts. I can't believe it's not lexical: Deriving distributed veridicality. In *Semantics and Linguistic Theory*, volume 29, pages 665–685, 2019.

[RU21] Floris Roelofsen and Wataru Uegaki. Searching for a universal constraint on the possible denotations of clause-embedding predicates. In *Semantics and Linguistic Theory*, volume 30, pages 542–561, 2021.

[SE15] Benjamin Spector and Paul Egré. A uniform semantics for embedded interrogatives: An answer, not necessarily the answer. *Synthese*, 192(6):1729–1784, 2015.

[TRA18] Nadine Theiler, Floris Roelofsen, and Maria Aloni. A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35(3):409–466, 2018.

[TRA+19] Nadine Theiler, Floris Roelofsen, Maria Aloni, et al. Picky predicates: Why believe doesn't like interrogative complements, and other puzzles. *Natural Language Semantics*, 27(2):95–134, 2019.

[Ueg19] Wataru Uegaki. The semantics of question-embedding predicates. *Language and Linguistics Compass*, 13(1), 2019.

[URR+22] Wataru Uegaki, Maribel Romero, Floris Roelofsen, et al. Mecore: Cross-linguistic investigation of meaning-driven combinatorial restrictions in clausal embedding, 2022. https://wuegaki.ppls.ed.ac.uk/mecore/ [Accessed: 21.02.2022].

[US19] Wataru Uegaki and Yasutada Sudo. The *hope-wh puzzle. *Natural Language Semantics*, 27(4):323–356, 2019.

[Whi21] Aaron Steven White. On believing and hoping whether. *Semantics and Pragmatics*, 14:6, 2021.

Modeling Metadisputes

Alexander W. Kocurek[1]

Cornell University

Philosophy, as well as life in general, is full of disputes that seem impossible to resolve. Sometimes, these disputes only *seem* impossible to resolve. One would hope that disputes over climate change, or the effectiveness of vaccines, are of this sort: were everyone to have all the facts, there would be no further residual disagreement. But other times, the disagreement would persist even with a consensus over the relevant worldly facts. Disagreements over whether a fetus is a person, or which logic is right, or whether free will is compatible with determinism, might be examples of this sort.

In light of an apparently irresolvable dispute D, we may adopt one of two views. First, there's **factualism**, which holds D is a factual dispute, i.e., there is a "fact of the matter" as to which side (if any) is correct. For factualists about D, the disagreement turns on a "substantive" matter in that the disputants have conflicting factual beliefs. Second, there's **semanticism**, which holds D is a verbal dispute, i.e., which side is correct is a "matter of interpretation".[1] For semanticists about D, the disagreement turns on a verbal matter in that the disputants (whether they realize it or not) are interpreting words differently.[2]

Let's say a **metadispute** is a dispute over whether another dispute is verbal or factual, i.e., a dispute between factualists and semanticists. My goal here is twofold. First, I aim to develop a formal theory of metadisputes that clarifies the differences between the factualist and semanticist positions and what's at stake. Second, I wish to motivate the thesis that metadisputes can themselves be verbal—a view I call **metasemanticism**. I will not present a full defense of this view here. Instead, I use the formal theory of metadisputes developed below to show that metasemanticism is at least a consistent view and briefly explain why one might think it is plausible of at least some prominent philosophical metadisputes.

A classic example of a metadispute comes from metaontology. Many ontological disputes—e.g., over whether composite objects exist, or whether objects persist via their temporal parts, or whether persons are physical or psychological entities—seem irresolvable. In light of this fact, some metaphysicians have held that ontological disputes are, in some sense, verbal (Carnap, 1950; Hirsch, 2005, 2009; Sidelle, 2007; Chalmers, 2009; Belleri, 2017; Thomasson, 2017). For example, Carnap (1950) famously claimed that metaphysical disputes turn on a choice of "framework". In a similar fashion, Hirsch (2005, 2009) holds that in many ontological disputes, each side can charitably interpret the other's language so that each side speaks the truth in their own language. More recently, Belleri (2017) and Thomasson (2017) argue ontological disputes are metalinguistic negotiations, i.e., normative metalinguistic disputes over how best to use language. Others have held these disputes are factual (Quine, 1948; Lowe, 1998; van Inwagen, 1998, 2009; Hawthorne, 2009; Sider, 2009, 2011). For example, Quine (1948) famously claimed ontological disputes are no more a linguistic matter than disputes within science. Following Quine, van Inwagen (1998, 2009) insists that being is univocal, not open to multiple interpretations. Sider (2009, 2011) holds that ontological disputes concern the most joint-carving interpretation of the existential quantifier (cf. Dorr 2005).

[1] The label 'semanticism' comes from Bennett (2009), though I will use this label less pejoratively in that, as I will discuss momentarily, a dispute may be "verbal" without being *merely* verbal, i.e., without involving speakers "just talking past each other". The label 'factualism', as far as I know, is my own.

[2] Factualism and semanticism are distinct, but closely related, to realism and antirealism respectively. How best to understand the relation between these views is a subtle and complex matter. For an overview in the context of metaontology, see Jenkins 2010.

These metaontological disputes turn on whether ontologists mean the same thing by the existential quantifier. Factualists hold that they do: the dispute concerns what exists according to some fixed privileged quantifier meaning. Semanticists hold they do not: each side is simply using 'exist' differently. Yet it has been difficult to settle whether ontologists do mean the same thing by the existential quantifier. Factualists point to the fact that ontologists often take themselves to be using the existential quantifier the same way. Principles of charity would thus seem to demand we take these convictions at face value even if that means one of them is mistaken about the disputed ontological claim. Semanticists, in response, argue principles of charity demand we take their ontological claims at face value even if that means ontologists are mistaken about whether they mean the same thing. Both factualists and semanticists appeal to principles of charity, though in different ways, and both postulate error somewhere in the dispute, though they locate the error in different places.[3] This raises the question: is there a fact of the matter as to which interpretation of ontological disputes is right? Or might these metadisputes *themselves* be verbal? And if so, how are we to make sense of this claim?

To understand such metadisputes, we must first understand what it means for a first-order dispute to be "verbal". By 'verbal disputes', I refer to **wholly verbal disputes**: disputes that solely arise from differences in interpretation (cf. Plunkett and Sundell 2013; Thomasson 2017). Some use the term to refer to **merely verbal disputes**: disputes where speakers "talk past each other", i.e., misunderstand what the other side is saying (cf. Chalmers 2011; Jenkins 2014). To illustrate, consider the following example (where the bracketed material is not pronounced):

(1) Ann: Hera called. She said she's here [in the city].
Ben: What are you talking about? Hera isn't here [in the house]!

Here, Ann and Ben have a merely verbal dispute. Ann and Ben do not *really* disagree over where Hera is: it simply *appears* that way to them due to a verbal confusion. But the subject matter of their apparent disagreement is a wholly factual one (viz., Hera's location).

By contrast, as I use the term, a "wholly verbal" dispute may involve a genuine disagreement over language. Consider a dispute between two astronomers, Ann and Ben:

(2) Ann: Pluto is clearly a planet. It has the right size, the right shape...
Ben: But what about Haumea? It has very similar properties.
Ann: I agree Haumea isn't a planet. But it's in the Kuiper belt.
Ben: So what? Surely we shouldn't define 'planet' so arbitrarily.
Ann: Why not? ...(etc.)

Here, Ann and Ben may know all the relevant astronomical facts, and may even know all the relevant facts about how the word 'planet' is used, including that each side is using it differently. Even so, a disagreement between them persists: it is not about how the term 'planet' *is* defined (whether by them or the relevant linguistic community), but rather about how it *should* be defined (i.e., a metalinguistic negotiation), or (as I'd prefer to put it) how *to* define it. They are trying to *decide* how to classify Pluto, not *discover* how Pluto is currently classified.

For the sake of space, I will restrict my attention in what follows to genuine disputes, i.e., disputes not based on a verbal confusion. So I will set aside merely verbal disputes and focus on debates over whether some dispute is wholly or partly verbal. Semanticism, then, is not committed to the claim that the dispute in question is pointless, or based on a misunderstanding: it just states that the dispute is over language.

[3] This is similar to the "speaker error objection" to metalinguistic negotiations: that interpreting a dispute as a metalinguistic negotiation requires postulating error on the part of the speakers as to what they're talking about. See Plunkett and Sundell 2021 for discussion and responses to this objection.

Elsewhere, I defend an account of wholly verbal first-order disputes as practical disputes over how to interpret words (Kocurek 2022; cf. Vermeulen 2018). I will briefly summarize the account before generalizing it to metadisputes. Following Grice (1968) (cf. Pinder 2021), distinguish **speaker meaning**, i.e., what a speaker means by an expression ("S means m by e"), from **semantic meaning**, i.e., what an expression means in a language or linguistic community ("e means m in L"). The former is largely determined by the speaker's intentions, whereas the latter by the collective intentions of a community/external factors (cf. Cappelen 2018). While speakers often intend to align their speaker meaning with semantic meaning, they may sometimes intend not to (e.g., to *advocate* for a new meaning). Thus, we should separate a speaker's **semantic beliefs**, i.e., their beliefs about what an expression semantically means, from their **semantic plans**, i.e., their plans for what speaker meaning to assign to an expression. Wholly verbal disputes arise when the speakers' semantic plans recommend different interpretations of the disputed claim given their beliefs.

To clarify this picture, let's model the meaning of an expression as an **intension**, i.e., a function from worlds to extensions. An **interpretation** is something that, at the very least, determines an assignment of intensions to every expression of the language. Here, we'll focus on the intensions assigned to sentences, i.e., propositions (generalization to subsentential expressions is straightforward). We can model a speaker's mental state S as a set of interpretation-world pairs (cf. Kocurek et al. 2020; Mena 2022).[4] Intuitively, $\langle i, w\rangle$ is in S if w is compatible with the speaker's beliefs and i is compatible with the speaker's semantic plans given that w obtains. We say S **accepts** a sentence ϕ if ϕ is true at every $\langle i, w\rangle \in S$. We say S **rejects** ϕ if ϕ is false at every $\langle i, w\rangle \in S$. Two speakers **disagree** over ϕ if one accepts ϕ while the other rejects it. A disagreement is **wholly verbal** if, roughly, they do not disagree over the proposition expressed by ϕ on any interpretation they leave open—or, put differently, their disagreement would dissolve were either side to adopt any interpretation left open by the other. Otherwise, it is **partly factual**. It is **wholly factual** if, roughly, they disagree over the proposition expressed by ϕ on any interpretation they leave open—or, put differently, their disagreement would persist were either side to adopt any interpretation left open by the other. Otherwise, it is **partly verbal**.[5]

On this account, metadisputes ultimately hinge on whether the two sides of the first-order dispute (speaker) mean the same thing by the disputed claim—or at least, whether there is enough of a shared (speaker) meaning that the dispute survives precisification. Yet whether this is so might itself be a matter of interpretation. One way this can happen is if we think beliefs and intentions come in degrees. On a widely held view of belief, belief is just a matter of having sufficiently high credence (see Jackson 2020 for an overview). But one could argue that what counts as "sufficiently high", even in a given context, is a matter of interpretation. Similarly, Goldstein (2016) defends a view of intention on which intention comes in degrees, arguing that full intention is just a matter of having a sufficiently high degree of intention. Thus, if semantic intentions come in degrees, then what a speaker means will depend on the threshold we set for full intention, which may be a matter of interpretation.

There is another, and perhaps more relevant, way to conclude that whether a dispute is verbal or factual can be a matter of interpretation: what a speaker means or believes may be

[4] Krifka (2012) develops a similar model of common ground in terms of a pair of a set of worlds and a set of interpretations. Here, we instead adopt the more fine-grained model using a set of pairs of interpretations and worlds. This helps capture conditional intentions ("If w obtains, interpretation e like so…").

[5] Since some disputes are partly verbal and partly factual, we should distinguish **(strong) factualism** about D, i.e., the view that D is *wholly* factual, from **weak factualism** about D, i.e., the view that D is *at least partly* factual—and likewise for semanticism. In principle, metadisputes may be between proponents of weak or strong versions of these views. To simplify, I'll restrict my attention to metadisputes between the strong versions.

underdetermined by the sentences they accept, as in cases of radical interpretation (Davidson, 1974; Dennett, 1987). As Davidson observes, there will likely be multiple ways of carving up a speaker's beliefs and intentions so that they come to accept the very same sentences. While he suggests that we might be able to narrow the interpretations of a speaker if we assume they are part of a community of speakers who all tend to coordinate their usage of language, he says, "it is not likely that only one theory will be found satisfactory." And if we leave open the possibility that the speakers are unknowingly engaged in a verbal dispute, this leaves more room for underdetermination.

Metasemanticism may therefore be plausible in cases where at least one of the speakers' beliefs or semantic plans are themselves open to interpretation. For example, a metasemanticist about ontological disputes would hold that even given all the sentences each side accepts, there are multiple ways to interpret the mental states of ontologists, some of which render their dispute factual and others which render it verbal. To illustrate how this could be, consider two ways to model the mental states of the mereological realist, who holds that composites exist, and the mereological nihilist, who does not. Let's stipulate, for ease of exposition, that both sides take realism and nihilism to be noncontingent: one of them is necessarily true and one of them is necessarily false. On a factualist model, both the realist and the nihilist interpret the existential quantifier in (roughly) the same way, but at least one of them (factually) believes something impossible. So if, say, composites are impossible, the realist's belief state can be modeled as a set of impossible worlds where composites exist. On a semanticist model, by contrast, both the realist and the nihilist have (roughly) the same factual beliefs, but their interpretation of the existential quantifier differs. So both the nihilist's and the universalist's beliefs may leave open the same possible worlds, but one interprets 'There are composite objects' as necessarily false whereas the other interprets it as possibly true (and, arguably, actually true). Put differently: the factualist models these mental states using the same interpretation but different state spaces, whereas the semanticist models them using different interpretations but the same state space.[6]

To be clear, I am not endorsing this line of reasoning. I present it just as a motivation for metasemanticism. There are ways to resist it, of course. For instance, one could insist that there are some objective criteria that rule out one of these models. But for present purposes, it does not matter whether such metadisputes are actually verbal. My goal is, first and foremost, to understand what it would mean to say they are.

We can model the difference between metafactualism and metasemanticism more formally using accessibility relations on interpretation-world pairs. Below, we work with a propositional language and restrict attention to disputes between two speakers over whether some claim ϕ is true. (Generalization to quantificational languages, disputes with more than two agents, and disputes over arbitrary questions is omitted.) Our language is defined as follows:

$$\phi ::= p \mid \neg\phi \mid (\phi \wedge \phi) \mid \langle \mathsf{f} \rangle \phi \mid \langle \mathsf{v} \rangle \phi$$

Here, $\langle \mathsf{f} \rangle \phi$ says the speakers at least partly factually disagree over ϕ and $\langle \mathsf{v} \rangle \phi$ says the speakers at least partly verbally disagree over ϕ. We can define other notions from these:

$$[\mathsf{f}]\phi := \langle \mathsf{f} \rangle \phi \wedge \neg \langle \mathsf{v} \rangle \phi$$
$$[\mathsf{v}]\phi := \langle \mathsf{v} \rangle \phi \wedge \neg \langle \mathsf{f} \rangle \phi$$
$$(\mathsf{d})\phi := \langle \mathsf{f} \rangle \phi \vee \langle \mathsf{v} \rangle \phi$$

[6] Notice the metasemanticist is not disputing the meaning of 'verbal dispute' or any other word. Rather, they are claiming it concerns how to interpret the dispute itself, i.e., how to interpret the mental states of the speakers involved. In this respect, the motivations for metasemanticism are not quite like those for first-order semanticism: the former is concerned with the interpretation of mental states, not with words directly.

Here, $[\mathsf{f}]\phi$ says the speakers have a wholly factual disagreement over ϕ, $[\mathsf{v}]\phi$ they have a wholly verbal disagreement over ϕ, and $(\mathsf{d})\phi$ they have some sort of disagreement over ϕ. Models are tuples of the form $\mathcal{M} = \langle W, I, S_0, S_1, [\![\cdot]\!]\rangle$ where W is a set of "worlds", I is a set of "interpretations", each $S_k \subseteq (I \times W)^2$ is an accessibility relation, and $[\![p]\!] \subseteq (I \times W)$.[7] The set $S_k(i,w) = \{\langle i', w'\rangle \mid \langle i, w\rangle S_k \langle i', w'\rangle\}$ represents the mental state of speaker k at w as interpreted by i. The set $B_k(i,w) = \{w' \mid \exists i' \colon \langle i', w'\rangle \in S_k(i,w)\}$ represents speaker k's belief state. The set $I_k(i,w) = \{i' \mid \exists w' \colon \langle i', w'\rangle \in S_k(i,w)\}$ represents the interpretations speaker k's semantic plans leave open given their beliefs. We define truth for arbitrary formulas as follows (dropping $\mathcal{M}$ for readability). For atomics, $[\![p]\!]^{i,w} = 1$ iff $\langle i, w\rangle \in [\![p]\!]$. Booleans are standard. For disagreement operators, where $[\![\phi]\!]^j = \{w \in W \mid \langle j, w\rangle \in [\![\phi]\!]\}$:

$$
\begin{aligned}
[\![\langle \mathsf{f}\rangle\phi]\!]^{i,w} = 1 \quad \Leftrightarrow \quad & \exists k \in \{0,1\} \colon\ S_k(i,w) \subseteq [\![\phi]\!] \text{ and } S_{1-k}(i,w) \subseteq [\![\neg\phi]\!] \text{ and either:} \\
& \text{(i) } \exists j \in I_{1-k}(i,w) \colon\ B_k(i,w) \subseteq [\![\phi]\!]^j \text{ or (ii) } \exists j \in I_k(i,w) \colon\ B_{1-k}(i,w) \subseteq [\![\neg\phi]\!]^j \\
[\![\langle \mathsf{v}\rangle\phi]\!]^{i,w} = 1 \quad \Leftrightarrow \quad & \exists k \in \{0,1\} \colon\ S_k(i,w) \subseteq [\![\phi]\!] \text{ and } S_{1-k}(i,w) \subseteq [\![\neg\phi]\!] \text{ and either:} \\
& \text{(i) } \exists j \in I_{1-k}(i,w) \colon\ B_k(i,w) \nsubseteq [\![\phi]\!]^j \text{ or (ii) } \exists j \in I_k(i,w) \colon\ B_{1-k}(i,w) \nsubseteq [\![\neg\phi]\!]^j .
\end{aligned}
$$

Roughly, assuming each speaker univocally adopts a single interpretation at the index of evaluation (i.e., their mental state is of the form $\{i\} \times B$), a disagreement over ϕ is at least partly factual iff they still disagree over the proposition expressed by ϕ on at least one of their interpretations. It is at least partly verbal iff they do not disagree over the proposition expressed by ϕ on at least one of their interpretations. More generally, a disagreement over ϕ is partly factual iff one of the speakers leaves open an interpretation of ϕ such that they still disagree even on that interpretation, whereas it is partly verbal iff one of them leaves open an interpretation of ϕ such that they do not disagree on that interpretation.

Figure 1 illustrates models of different kinds of "first-order" disputes in this framework. Each node represents an interpretation-world pair (the interpretations determine the "rows" while the worlds determine the "columns"). The arrows represent the accessibility relations of the two speakers. In each case, the blue speaker (S_1) accepts p at each node while the red speaker (S_0) rejects p—in other words, for each $\langle i, w\rangle$ pair in the model, $S_1(i,w) \subseteq [\![p]\!]$ while $S_0(i,w) \subseteq [\![\neg p]\!]$.[8] To check whether the speakers disagree over ϕ (i.e., whether $(\mathsf{d})\phi$ is true) at a given node, simply check that every blue-accessible node, and no red-accessible node, satisfies ϕ. One can determine the status of a disagreement over ϕ as follows. First, vertically move the blue shading on the blue-accessible nodes to each row with a red-accessible node, and check whether ϕ is false on any of those nodes. Next, do the same for the red shading on the red-accessible nodes, checking whether ϕ is true on any of those nodes. If the answer is always no, then the disagreement is wholly factual. If the answer is always yes, then the disagreement is wholly verbal. If the answer is sometimes yes and sometimes no, then it is partly both.

In these cases, no metadispute arises: the speakers involved agree over the status of their dispute. Figure 2 illustrates cases where the speakers also have metadisputes over the status of their dispute. In both cases, the blue speaker accepts their dispute over p is wholly factual while the red speaker accepts it is wholly verbal. In the leftmost model, the speakers agree their metadispute is wholly factual. In the rightmost model, the blue speaker accepts their metadispute is wholly factual while the red speaker accepts it is wholly verbal; and similarly for their *meta*metadispute, etc. This shows that it is not only consistent for one to accept either metasemanticism, but it is even consistent for one to accept semanticism "all the way up".

[7] Interpretations (i.e., members of I) are treated as arbitrary points. Their associated assignment of intensions is determined by $[\![\cdot]\!]$.

[8] Note, this includes the uncolored nodes in the rightmost model, though only vacuously so, since $S_k(i_2, w_1) = S_k(i_1, w_2) = \varnothing$ for each k.

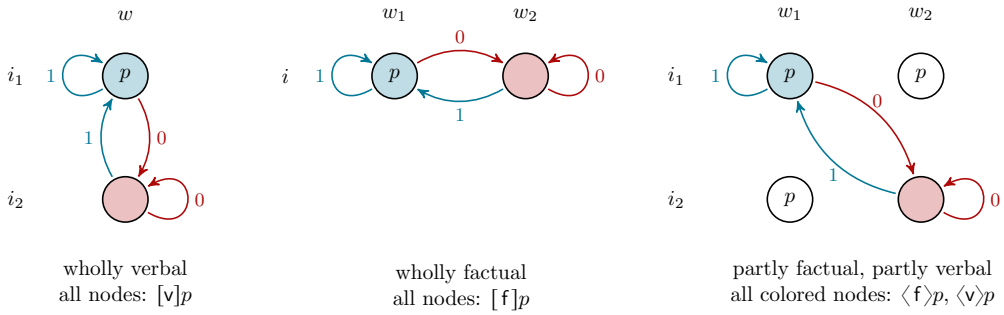


Figure 1: Models of different kinds of disputes. Nodes represent interpretation-world pairs. Arrows represent the accessibility relation. In each case, the dispute has the same status at every colored interpretation-world pair.

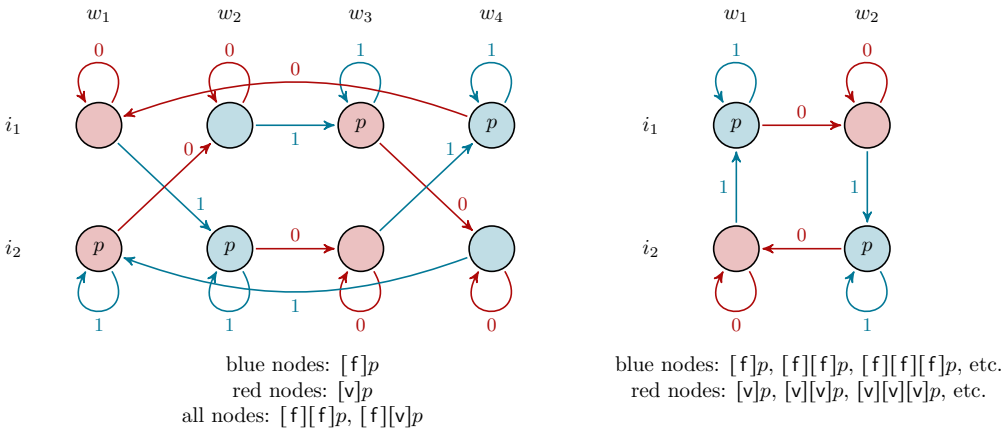


Figure 2: Models of different kinds of metadisputes. In the left model, the speakers disagree over p and disagree over whether their dispute is verbal but agree their metadispute is factual (metafactualism). In the right model, the speakers disagree over p, over whether their dispute is verbal, over whether their metadispute is verbal, etc. The blue speaker accepts factualism at all orders while the red speaker accepts semanticism at all orders.

In closing, return to the metaontology example. As noted earlier, metadisputes over ontological disputes turn on whether ontologists mean the same thing by the existential quantifier. So if we think there is no fact of the matter as to whether we should model the mental states of ontologists in terms of shared beliefs or shared semantic plans, the dispute over whether ontological disputes are verbal may, itself, be verbal. Not only is such a view technically possible, but also there are at least some reasons for thinking in could be true: there are factualist and semanticist models of the mental states of ontologists and it is hard to see what objective fact would settle which is correct. All this suggests that metasemanticism is a viable *meta*metaontological view worth taking seriously. Whether it is, as I suspect, the most promising such view will have to be taken up elsewhere.

References

Belleri, Delia. 2017. "Verbalism and metalinguistic negotiation in ontological disputes." *Philosophical Studies* 174:2211–2226.

Bennett, Karen. 2009. "Composition, Colocation, and Metaontology." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 38–76. Oxford: Oxford University Press.

Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.

Carnap, Rudolf. 1950. "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 4:20–40.

Chalmers, David J. 2009. "Ontological Anti-Realism." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 77–129. Oxford: Oxford University Press.

—. 2011. "Verbal Disputes." *The Philosophical Review* 120:515–566.

Davidson, Donald. 1974. "On the Very Idea of a Conceptual Scheme." *Proceedings and Addresses of the American Philosophical Association* 47:5–20.

Dennett, Daniel C. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Dorr, Cian. 2005. "What we disagree about when we disagree about ontology." In Mark E. Kalderon (ed.), *Fictionalist Approaches to Metaphysics*, 234–286. Oxford: Oxford University Press.

Goldstein, Simon. 2016. "A Preface Paradox for Intention." *Philosophers' Imprint* 16:1–20.

Grice, Paul H. 1968. "Utterer's meaning, sentence-meaning, and word-meaning." *Foundations of Language* 4:225–242.

Hawthorne, John. 2009. "Superficialism about Ontology." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, chapter 7, 213–230. Oxford University Press.

Hirsch, Eli. 2005. "Physical-Object Ontology, Verbal Disputes, and Common Sense." *Philosophy and Phenomenological Research* 70:67–97.

—. 2009. "Ontology and Alternative Languages." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 231–259. Oxford University Press.

Jackson, Elizabeth G. 2020. "The relationship between belief and credence." *Philosophy Compass* 15:e12668.

Jenkins, C. S. 2010. "What Is Ontological Realism?" *Philosophy Compass* 5:880–890.

—. 2014. "Merely Verbal Disputes." *Erkenntnis* 79:11–30.

Kocurek, Alexander W. 2022. "Verbal Disagreement and Semantic Plans." Submitted.

Kocurek, Alexander W., Jerzak, Ethan J., and Rudolph, Rachel E. 2020. "Against Conventional Wisdom." *Philosophers' Imprint* 20:1–27.

Krifka, Manfred. 2012. "Definitional Generics." In Alda Mari, Claire Beyssade, and Fabio del Prete (eds.), *Genericity*, 372–389. Oxford: Oxford University Press.

Lowe, E J. 1998. *The Possibility of Metaphysics: Substance, Identity, and Time*. Oxford: Oxford University Press.

Mena, Ricardo. 2022. "Metalinguistic effects." *Mind & Language* 1–21. doi:10.1111/mila.12390.

Pinder, Mark. 2021. "Conceptual Engineering, Metasemantic Externalism and Speaker-Meaning." *Mind* 130:141–163.

Plunkett, David and Sundell, Tim. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13:1–37.

—. 2021. "Metalinguistic negotiation and speaker error." *Inquiry* 64:142–167.

Quine, Willard van Orman. 1948. "On What There Is." *The Review of Metaphysics* 2:21–38.

Sidelle, Alan. 2007. "The Method of Verbal Disputes." *Philosophical Topics* 35:83–113.

Sider, Theodore. 2009. "Ontological Realism." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 384–423. Oxford: Oxford University Press.

—. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Thomasson, Amie L. 2017. "Metaphysical Disputes and Metalinguistic Negotiation." *Analytic Philosophy* 58:1–28.

van Inwagen, Peter. 1998. "Modal Epistemology." *Philosophical Studies* 92:67–84.

—. 2009. "Being, Existence, and Ontological Commitment." In David J Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*, 472–506. Oxford: Oxford University Press.

Vermeulen, Inga. 2018. "Verbal Disputes and the Varieties of Verbalness." *Erkenntnis* 83:331–348.

The Role of Questions, Circumstances, and Algorithms in Belief

Jens Kipper[1], Alexander W. Kocurek[2], and Zeynep Soysal[1]

[1] University of Rochester
[2] Cornell University

A recent approach to the problem of logical omniscience holds that belief is *question-sensitive*: belief is not a two-place relation between an agent and a proposition, but rather a three-place relation between an agent, a question, and an answer to that question (Pérez Carballo, 2016; Yalcin, 2018; Hoek, 2022). While the question-sensitive approach can avoid some logical omniscience problems, we argue that it suffers from nearby problems. Recognizing the flaws of this approach, however, naturally leads to a more promising solution.

On the classical picture, a belief state is modeled as set of possible worlds, viz., those worlds compatible with what the agent believes (Hintikka, 1962; Stalnaker, 1984).[1] An agent believes a proposition P iff their belief state B entails P (i.e., $B \subseteq P$). Commonly, this picture is combined with the idea that there are constitutive links between belief and (rational) action. An agent who believes P is disposed to act on P, which means, roughly, that they are disposed to bring about desire-satisfaction in worlds in which P and the agent's other beliefs are true (see Stalnaker 1984). This picture of belief has many advantages (Stalnaker, 1984; Yalcin, 2018). For one, it explains the holistic nature of belief update. If an agent comes to believe they left their keys on the table, they will often automatically come to believe their keys are on the table, their keys are not in their pocket, and so on. It also explains how we can attribute to agents beliefs in propositions they haven't considered. If a driver sees a moose in the road and swerves to avoid a collision, it is fair to say they believe the moose is larger than a golf ball, even if they have not explicitly articulated this belief.

Despite these advantages, however, the classical picture faces the well-known *problem of logical omniscience:* it incorrectly predicts that belief is closed under entailment. For if a belief state B entails P and P entails Q, then B entails Q. So for example, on the classical picture, since 107 is necessarily prime, if Chip believes the CVV of his debit card is 107, he must also believe the CVV is prime. Yet it is clear that Chip can have the former belief without the latter. As a special case, the classical picture predicts that an agent always believes any necessary truth and that their beliefs are closed under equivalence. So, continuing the previous example, the classical account predicts that Chip believes 107 is prime and that if he believes his CVV is 107, then he believes his CVV is the 28th prime number.

The question-sensitive account of belief is designed to retain the core benefits of the classical account while allowing for failures of closure. On this picture, agents do not simply believe propositions: they believe *answers to questions*. Questions effectively determine which propositions are "visible" to an agent: relative to a question, agents can only believe propositions that are answers to that question. So, while Chip may believe his CVV is 107 relative to the question *What's the CVV?*, he may not believe this relative to the question *What information does my debit card contain?* (e.g., he may not be able to recall his debit card includes the CVV, even though he could recite the CVV if asked about it explicitly).

[1] To account for egocentric belief, i.e., belief about one's location in logical space, belief states can also be modeled as sets of *centered* worlds, where a centered world is a pair of a possible world and an individual at a time—see Lewis 1979. While we find this refinement of the classical picture useful, we still construe the worlds involved as uncentered possible worlds here, for the sake of simplicity.

Most defenders of this account follow the early literature on the semantics of questions, which model questions as partitions on possible worlds (Hamblin, 1973; Groenendijk and Stokhof, 1984; Lewis, 1988). A belief state is thus a set of *pointed partitions*, i.e., pairs $\langle Q, A\rangle$ where Q is a partition on possible worlds (a question) and A is a union of Q-cells (an answer to Q).[2] An agent believes A relative to Q iff (i) A is an answer to Q, and (ii) there is a pointed partition $\langle Q, B\rangle$ in their belief state such that $B \subseteq A$. This account retains the main benefits of the classical account (e.g., the holistic nature of belief update) while allowing for failures of closure. For example, Chip may believe the CVV is 107 relative to *What's the CVV?* without believing the CVV is prime relative to *Is the CVV prime?*, since these are different questions.[3]

There are two main problems with such accounts. First, they all validate the following implausible closure principle: if A entails B, both of which are answers to Q, then if an agent believes A relative to Q, they also believe B relative to Q.[4] For example, the proposition *The CVV is prime* is an answer (albeit a partial one) to *What's the CVV?*, and it is entailed by *The CVV is 107*. Hence, these accounts implausibly predict that if Chip believes the CVV is 107 relative to *What's the CVV?*, then he also believes the CVV is prime relative to the same question.[5] This closure principle also entails two weaker principles, both of which are problematic. The first involves necessary answers: if an agent has any belief relative to Q, then they believe every necessarily true answer to Q. Thus, if Chip believes 107 is either prime or composite relative to *Is 107 prime?*, then he also believes 107 is prime relative to this question. The second problematic principle involves equivalence: if an agent believes A relative to Q, where A is necessarily equivalent to B and Q is necessarily equivalent to R (meaning a complete answer to one is a complete answer to the other), then they believe B relative to R. Thus, if Chip believes his CVV is among 101, 103, 107, ... (listing all the 3-digit primes) relative to *Is the CVV among 101, 103, 107, ...?*, then he also believes his CVV is a 3-digit prime relative to *Is the CVV a 3-digit prime?*. These predictions seem as problematic as the predictions of the classical account.

Second, question-sensitivity by itself is not suitable for explaining many kinds of failures of logical omniscience. Question-sensitive accounts are supposed to explain failures of logical omniscience in terms of limitations of computational resources. For instance, Yalcin (2018) argues that failures to recognize the logical consequences of our beliefs are due to our inability to resolve logical space in full detail.[6] A partition represents how finely an agent resolves logical space. Finer resolutions reduce logical ignorance but come with higher "encoding costs" in the form of *information potential*—roughly, the number of propositions needed to construct the partition, which is just the logarithm (base 2) of the number of cells in the partition (bits) rounded up to the nearest integer. Thus, failures of logical omniscience are connected to an agent's lack of resources for representing very fine-grained partitions.

[2] Some authors impose further constraints on belief states. See, e.g., Yalcin 2021; Hoek 2022. Our criticisms of question-sensitive accounts is independent of such constraints.

[3] For a question-sensitive account of the link between belief and action, see Hoek (2022).

[4] Yalcin (2018) refers to this principle (for "accessible belief") as closure under "visible consequence". While he acknowledges this principle may be undesirable, his reasons are different: his worry is that it renders too many propositions visible to the agent. Even if we set aside these sorts of worries, however, we think this closure principle has undesirable consequences.

[5] Teague (forthcoming) gives two counterexamples to a stronger principle: if every complete answer to Q entails a complete answer to R, and an answer A to Q entails an answer B to R, then if an agent believes A relative to Q, they believe B relative to R. However, while some accounts validate this principle (Hoek, 2022, forthcoming), others do not (Yalcin, 2018, 2021). Moreover, Teague's counterexamples rely on lacking concepts and lacking attention. By contrast, our counterexample does not: Chip may possess the concept PRIME and be attending to the question of whether the CVV is prime yet still fail to believe it is. Our counterexample is thus more pressing for question-sensitive accounts.

[6] See Pérez Carballo (2016) and Hoek (forthcoming) for similar ideas.

However, there are other, and plausibly more relevant, sources of computational costs associated with determining logical consequences. Consider, e.g., the game of chess. The rules of chess determine, for any given position, whether there is a forced checkmate. Moreover, a competent player can resolve any chess position in full detail. They may even have an algorithm, such as minimax, that can decide whether there is a forced checkmate in any given position. Nevertheless, in many positions, this algorithm would take much too long to conclude, and thus chess players often fail to know (or have a belief about) whether there is a forced checkmate. By contrast, it is much easier to determine the position of the chess board. Even though the information potential of *Is there a forced checkmate?* is small (1 bit) compared to *What's the position?* (≈133 bits), the relevant computational costs of the former relate to time-complexity, not "encoding costs". Moreover, there are other kinds of failures of logical omniscience that aren't captured by the question-sensitive approach. One comes from how a question is presented to an agent. For example, Chip may be able to recall his CVV as an answer to a multiple-choice question (with a "None of the above" option) but unable to recall it as an answer to a fill-in-the-blank question. Others come from an agent's psychological state, or from the external circumstances they are in. For instance, Chip may be able to recall his CVV in normal conditions but not when he is stressed or when loud music is playing in the background.

The account we suggest is designed to encompass these other sources of failures of logical omniscience as well. Like both the classical account and question-sensitive accounts, our account is based on the idea that an agent who believes ϕ is disposed to act on the information associated with ϕ. We use two implications of this idea.

First, as was explained above, a disposition to act on (the information associated with) ϕ can be understood, roughly, as one to bring about desire-satisfaction in worlds where it is true. Here, desire-satisfaction is the manifestation condition of this disposition and (the information associated with) ϕ is its trigger condition. Now, dispositions are generally dependent on *ceteris paribus* or normal conditions. These are the conditions that need to be in place for a disposition to be manifested, if the trigger occurs. For instance, a normal condition for a match to manifest its disposition to light if struck is the presence of oxygen. The same applies to doxastic dispositions. Accordingly, belief is not just *question*-sensitive but, more generally, *circumstance*-dependent, i.e., dependent on the set of normal conditions being considered. This implies that for an agent to be able to satisfy their desires, they not only need to have true beliefs, but the relevant circumstances must obtain as well. These circumstances can concern not just the question an agent faces, but anything about their cognitive or physiological functioning, or their external conditions.[7]

[7]In a similar vein, Elga and Rayo (2021) characterize the mental states of agents using "conditons". They develop a model of fragmented agents in terms of "access tables", which map conditions to credence functions, and then postulate that it is appropriate to attribute a full belief to a fragmented agent iff the dispositions predicted by that belief are sufficiently similar to those predicted by the access table. However, our approach is different in several respects. First, Elga and Rayo use conditions to characterize fragmentation: different conditions correspond to different fragments of an agent's mental state. By contrast, our use of circumstances is not for characterizing fragments but rather for characterizing the dispositions associated with a belief. Our model of belief does not require agents to be fragmented to avoid logical omniscience. Second, while Elga and Rayo reduce full belief to dispositions to act, they do not attempt to formally represent the dispositions associated with a given belief, whereas our model does. Third, Elga and Rayo postulate an "obviousness function" associated with each condition that determines (roughly) which inferences are "obvious" in that fragment and which are not. Our account instead appeals directly to an agent's dispositions (and their decision-making algorithms) to determine which inferences are they find "obvious" in a given circumstance. Lastly, on their account, each condition is associated with a normal credence function, which means that the information accessible relative to given condition is closed under entailment. Thus, their account validates something akin to the question-sensitive closure principle we reject.

Second, it is highly plausible that an agent's dispositions to act are produced by (internal) algorithms. Algorithms can thus play an important explanatory role in an account of belief: whether one acts on a piece of information in certain circumstances depends on the algorithm one is disposed to deploy in these circumstances.[8] On our view, then, believing that ϕ means being disposed to employ an algorithm which outputs acting on the information associated with ϕ in normal circumstances associated with ϕ. Typically, this means that the algorithm outputs the relevant action in a relatively short amount of time: since most of the decisions we face are time-sensitive, an agent whose algorithms are very slow will usually not be able to act in ways that bring about desire-satisfaction.[9]

We formalize this account as follows. Given a background language $\mathcal{L}$ and a set of worlds W, we capture circumstance-dependency by assuming that each $\phi \in \mathcal{L}$ at each $w \in W$ is associated with a set $C_{\phi,w} \subseteq W$ of "normal" worlds for ϕ at w.[10] Let $[\phi]$ be the set of worlds where ϕ is true. To capture algorithm-dependency, we first model decision problems as pairs $\langle O, t\rangle$, where O is a set of options or choices the agent can make and t is the amount of time an agent has to make their choice. Let Δ_w be the set of decision problems an agent faces in w. Then at each world $w \in W$, an agent is associated with a decision algorithm α_w that takes decision problems in Δ_w as inputs and outputs an option after a certain amount of time if it halts. We write $\alpha_w(O, t) = a$ if α outputs $a \in O$ by t; otherwise, $\alpha_w(O, t)$ is the "null-action" (do nothing).

There are many ways to flesh out the notion of "acting on" some information. For concreteness, we assume agents "act" on information by avoiding options that would be weakly dominated were the information accurate (see Hoek 2022). This is a natural way of making precise the idea that acting on the information that ϕ means acting in ways that bring about desire-satisfaction in ϕ-worlds. More precisely, where $P \subseteq W$, an option $a \in O$ is weakly P-dominated by $b \in O$ if the utility an agent receives by performing a is no greater, and in some cases strictly less, than the utility they receive by choosing b in any P-world. An agent acts on P, then, if they avoid weakly P-dominated options.

Now, it seems plausible that an agent can be associated with a belief that ϕ without having a disposition to act on $[\phi]$ itself. To take an extreme case, if $[\phi] = \varnothing$ (i.e., ϕ is impossible), then no option weakly $[\phi]$-dominates any other. Thus, if we allow belief in impossible propositions, the dispositions to act on them cannot just be the disposition to act on $[\phi]$ itself. However, even for contingent propositions, having the disposition to act on *that* proposition is plausibly not the same as believing it. Suppose the bank teller asks Chip for his CVV to verify his identity. Arguably, if Chip believes his CVV is 107, he should be disposed to answer "107" in this sort of case. But if the belief that his CVV is 107 is only associated with the disposition to act on [The CVV is 107], then we will fail to predict this disposition. For there are [The CVV is 107]-worlds where answering "107" does not bring about desire-satisfaction for Chip—e.g., the bank teller might be trying to steal his debit card information to commit fraud, and in those worlds, answering with a random false number would be preferable. Indeed, there are relatively few decision problems that have weakly dominated options at *all* [The CVV is 107]-worlds. But, intuitively, many of those worlds do not seem relevant for characterizing the information Chip is disposed to act on.

One way to resolve this issue is to say that a belief ascription is not associated with a disposition to act on the content of the ascribed belief, but rather to act on some related proposition.

[8] Algorithmic accounts of belief and knowledge have been developed by Parikh (1987); Halpern et al. (1994). See Soysal 2022 for a recent algorithmic account that is tied to a possible-worlds account of content.

[9] See Stalnaker 1991, 437 for discussion of this issue.

[10] Plausibly, it would be more adequate to construe circumstances in terms of centered worlds, accounting for the idea that these represent the circumstances an agent may find themselves in (see footnote 1.) Since this issue isn't relevant to our discussion, we stick with uncentered worlds here, for the sake of simplicity.

For the sake of generality, let us introduce a contextually-supplied function $(\cdot)$, which takes formulas as inputs and outputs propositions (cf. Williamson's (2009) use of "counterparts" of formulas for knowledge ascriptions). We can think of (ϕ) as the "actionable content" of ϕ for the agent, i.e., the relevant contextual information that an agent is disposed to act on given that it is correct to ascribe to them a belief that ϕ. For example, in the case of Chip, (The CVV is 107) may be the conjunction of the proposition that the CVV is 107 together with other contextual information. If $[\phi] = \varnothing$, then (ϕ) might be a metalinguistic proposition that the sentence ϕ is true (perhaps conjoined with relevant contextual information). Then our account can be stated as follows: an agent believes ϕ at w iff for all $v \in C_{\phi,w}$, all $\langle O, t\rangle \in \Delta_v$, and all $a \in O$, if a is weakly (ϕ)-dominated by some $b \in O$, then $\alpha_v(O, t) \neq a$.

This account is suitable for solving the problem of logical omniscience, since it does not entail any of the problematic closure conditions that the classical account or the question-sensitive accounts are committed to. Due to the fact that we construe belief as relative to circumstances, an agent's believing ϕ only entails that they believe ψ if it is both the case that $(\phi) \subseteq (\psi)$, i.e., the actionable content of ϕ contains that of ψ, and that $C_{\psi,w} \subseteq C_{\phi,w}$. The second conjunct ensures that if someone can bring about desire-satisfaction in all circumstances associated with ϕ, they can do so in all those associated with ψ. To see how this account avoids closure under entailment, assume that $(\phi) \subseteq (\psi)$ and that $C_{\psi,w} \nsubseteq C_{\phi,w}$. Then an agent who is disposed to employ an algorithm that outputs acting on the information that (ϕ) in $C_{\phi,w}$ but not in $C_{\psi,w}$ believes that ϕ without believing that ψ.

As an illustration, it might be part of the circumstances associated with a belief which task one is facing. For instance, to believe that the CVV is among 101, 103, 107, ..., one plausibly needs to have an algorithm that quickly outputs "Yes" if one is (verbally) asked "Is the CVV among 101, 103, 107, ...?" but not one that quickly outputs "Yes" if one is asked "Is the CVV a three-digit prime?". Since an agent can have the former kind of algorithm without having the latter, this explains why one can believe one without believing the other, despite the fact that the representational content of the two beliefs is the same, and despite the fact that the questions the agent are asked are equivalent.

This example is also suggestive of how one can account for failures to have beliefs whose contents are necessarily true. If (ϕ) is necessarily true, then the only reason why an agent can fail to believe ϕ is because they do not have an algorithm that outputs desire-satisfaction in $C_{\phi,w}$. Assume, e.g.—as seems plausible—that to believe that 107 is prime, one needs to be able to respond "Yes" in some limited amount of time if one is asked "Is 107 prime?", given that one desires to give the correct answer. On our account, this means that this task is among the circumstances associated with "107 is prime". Hence, since an agent can fail to have an algorithm that outputs the correct response sufficiently quickly, an agent can fail to believe that 107 is prime.

While we reject commonly held closure conditions on belief, our account can nevertheless capture the fact that beliefs often come in packages. For instance, an agent who believes that they just saw a moose larger than a golf ball will typically also believe that they just saw an animal, that they just saw an object larger than a bottle cap, etc. On our account, this can be explained by appealing to the algorithms involved. Often, one algorithm available to an agent grounds many of their dispositions to act and, thus, confers many beliefs to them. Furthermore, many beliefs are very commonly (but not necessarily) conferred by specific algorithms. Taken together, this explains why an agent with a particular belief can often be expected to have many other beliefs—due to the fact that the algorithm that confers one of these beliefs also confers the others. For instance, an algorithm that produces desire-satisfaction with respect to questions about whether some number is divisible by 6 usually also produces desire-satisfaction

with respect to questions about whether some number is divisible by 2, because the standard method most people use to check divisibility by 6 involves checking for divisibility by 2. This can explain why someone who believes, say, that 294 is divisible by 6 will typically also believe that 294 is divisible by 2. But they won't necessarily believe the latter, for they might use alternative methods for answering questions about divisibility by 6, such as having memorized all numbers divisible by 6 up to 294. The holistic nature of belief update, mentioned above, can be explained in a similar manner: since possession of one algorithm often confers many different beliefs, we can expect an agent who acquires a new belief to automatically acquire many other beliefs. Finally, similar types of constraints on algorithms could be used to capture rational or competent belief. For instance, one might construe logically competent agent as those who employ algorithms that "respect" the standard inference rules for the logical connectives, in the sense that their algorithm computes, say, a conjunction to be true just in case it computes its conjuncts to be true.

The involvement of algorithms also allows us to explain many types of failure of logical omniscience. Very commonly, if an agent fails to believe a logical consequence of something they believe, this is either because they do not have an algorithm that allows them to derive this consequence, or because they only have an algorithm that would take too many computational resources to derive it. Take again Chip, who believes that his CVV is among 101, 103, 107, ... but does not believe that it is prime. The latter might be because he does not have an algorithm that can determine primality, or, more likely, because he does have such an algorithm, but it would take too long for him to run it. Either way, Chip is unable to give the correct answer to "Is the CVV prime?" in a reasonable amount of time and hence, he lacks the relevant belief.

In closing, let us return to the question suggested in the title: what role do questions, circumstances, and algorithms play in belief? We argued that question-sensitive accounts are limited in scope and only solve restricted versions of the problem of logical omniscience while leaving the full version unresolved. Even so, we argued that question-sensitive accounts are picking up on a specific instance of a more general feature of belief, viz., circumstance-dependency. By modeling belief as sensitive to both ceteris paribus conditions and the algorithms an agent uses to act, we developed a more robust and viable account of belief that solves the problem of logical omniscience in full generality.

Our account of belief leaves open some unanswered questions that we leave to future research. First, our discussion here has focused on full belief, and thus we have left it open how to generalize our account to partial belief or credences. Second, it might be useful to specify the notion of "actionable content" further by introducing constraints on the $(\cdot)$ function. Such constraints have ramifications for the logic of belief, the details of which could be examined in future work.[11] Lastly, as we explained above, our account reduces the rationality of belief to constraints on the decision-making algorithms an agent employs. We suggested one constraint we might impose on logically competent agents, i.e., agents who understand the logical connectives. We can obtain more idealized forms of rationality by imposing stronger constraints, e.g., requiring an agent's algorithm to avoid weakly dominated actions in *all* circumstances. We leave it open here how other aspects of rationality can be captured by introducing constraints on algorithms.

[11] One constraint worth considering is the following: if $[\phi] \subseteq [\psi]$, then $(\phi) \subseteq (\psi)$. This ensures that if ϕ entails ψ, then being disposed to act on ϕ in certain circumstances entails being disposed to act on ψ *in the same circumstances*. (Note: this does not imply that an agent who believes ϕ also believes ψ since the set of circumstances associated with ϕ and ψ might still differ.) Another plausible constraint: where T is a truth predicate, $(\phi) \subseteq [\phi] \cup [T(\ulcorner\phi\urcorner)]$. This would ensure that if $[\phi] = \varnothing$, then $(\phi) \subseteq [T(\ulcorner\phi\urcorner)]$, i.e., (ϕ) at least contains the metalinguistic information that the sentence ϕ is true.

References

Elga, Adam and Rayo, Agustín. 2021. "Fragmentation and Logical Omniscience." *Noûs* 1–26. doi:10.1111/nous.12381.

Groenendijk, Jeroen and Stokhof, Martin. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam, Department of Philosophy.

Halpern, Joseph Y., Moses, Yoram, and Vardi, Moshe Y. 1994. "Algorithmic Knowledge." In *Proceedings of the 5th Conference on Theoretical Aspects of Reasoning and Knowledge (TARK'94)*, 255–266. Morgan Kaufmann.

Hamblin, Charles Leonard. 1973. "Questions in Montague English." *Foundations of Language* 10:41–53.

Hintikka, Jaakko. 1962. *Knowledge and Belief: An Introduction to the Logic of Two Notions*. Ithaca: Cornell University Press.

Hoek, Daniel. 2022. "Questions in Action." *Journal of Philosophy* 1–23.

—. forthcoming. "Minimal Rationality and the Web of Questions." In Dirk Kindermann, Peter van Elswyk, and Andy Egan (eds.), *Unstructured Content*, 1–30. Oxford: Oxford University Press.

Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 513–543.

Lewis, David K. 1988. "Statements Partly About Observation." *Philosophical Papers* 17:1–31.

Parikh, Rohit. 1987. "Knowledge and the Problem of Logical Omniscience." In Z. W. Ras and M. Zemankova (eds.), *International Symposium on Methodologies of Intelligent Systems*, 432–439. North-Holland.

Pérez Carballo, Alejandro. 2016. "Structuring Logical Space." *Philosophy and Phenomenological Research* 92:460–491.

Soysal, Zeynep. 2022. "A Metalinguistic and Computational Approach to the Problem of Mathematical Omniscience." *Philosophy and Phenomenological Research* 1–20. doi:10.1111/phpr.12864.

Stalnaker, Robert C. 1984. *Inquiry*. Cambridge, MA: MIT Press.

—. 1991. "The Problem of Logical Omniscience, I." *Synthese* 89.

Teague, Richard. forthcoming. "The Problem of Closure and Questioning Attitudes." *Synthese* 1–24.

Williamson, Timothy. 2009. "Probability and Danger." *The Amherst Lecture in Philosophy* 4:1–35.

Yalcin, Seth. 2018. "Belief as Question-Sensitive." *Philosophy and Phenomenological Research* 97:23–47.

—. 2021. "Fragmented but Rational." In Cristina Borgoni, Dirk Kindermann, and Andrea Onofri (eds.), *The Fragmented Mind*, chapter 6, 156–179. Oxford: Oxford University Press.

Alternate Categorization: a Formal-Conceptual Semantics of Reciprocal Alternations *

Imke Kruitwagen[1], James A. Hampton[2], Yoad Winter[1], and Joost Zwarts[1]

[1] Utrecht University
[2] City, University of London

1 Introduction

The reciprocal alternation is a cross-linguistically common pattern of variation between verbal forms and their meanings [5]. It involves collective predicates (1a) and/or 'with' forms (1b) that alternate with binary predicates: transitive verbs (1c) and verbs with prepositional complements (1d).

(1) a. Sue and Dan fought/talked.
b. Sue fought/talked with Dan.
c. Sue fought Dan.
d. Sue talked to Dan.

While the semantics of alternations as in (1) has been extensively discussed in the literature, there is currently no general account of the intricate semantic relations between these forms. The goal of this paper is twofold: to present a formal model of these semantic relations, and to examine enough data that allow us to evaluate it. In an earlier work (Kruitwagen et al. [4]), we presented a prototype model of collective sentences like (1a). Here we extend this model into a general semantic threshold-based model. In this model, a verbal root has a *conceptual core* specifying the attributes that underly meanings of different verbal forms, although the weights of those attributes vary between alternations. The model is tested against old data, and new data from five different verbs showing the 'with' alternation (1b) and other binary forms (1c-d). The results show strong support for our proposed model, which unlike previous alternatives, accounts for (the lack of) entailments between alternates as in (1) without compromising the formal semantic coherence of the theory. The result is a semantic theory that is flexible enough, although restricted within boundaries of the root's conceptual core.

2 The theoretical proposal

In many previous works (see [1, 8], among others) it is proposed that eventive reciprocal sentences (1a) are logically related to their 'with' alternates (1b). Both kinds of sentences are assumed to involve full symmetry between the participants of the event. When there are only two participants, this approach entails that collective sentences (1a) are equivalent to their 'with' alternates (1b), as well as to a conjunction of binary predications like *Sue fought Dan and Dan fought Sue* describing one symmetric event. Rákosi [6] challenges the first equivalence, arguing that 'with' sentences like *A fought with B* require a lesser extent of agency from B

*The research was supported by a grant of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 742204).

than from A. On the basis of Hungarian data about such sentences, Rákosi postulates a new thematic label ('partner') for 'with' arguments. However, the semantic properties of this role and its differences from classic 'agent' and 'patient' remain unanalyzed.

Our proposal relies on the common assumption that the abstract morphology and semantics of a verbal root feed the verb's different forms [7]. Semantically, we propose that each root r denotes a complex *conceptual core* (CC). The CC is an aggregate of primitive features that make up the root's general meaning. Of these features we focus on *activity* (A_r) and *intentionality* (I_r) of the root's meaning. For instance, for the root *fight* (f) we have A_f='physically oppose' and I_f='be hostile to'. These features are treated as functions from events and pairs of entities to the real interval $[0, 1]$. The A_f and I_f levels that an entity x reaches towards an entity y in an event e determine the typicality of that event as an instance where x fights y. That typicality makes the sentence true if it exceeds a pragmatic *threshold* [3]. Formally, a binary form (e.g. transitive *fight, talk to*) of a root r denotes the 3-place predicate, which is defined in terms of the CC features and the threshold as follows:

$$\mathbf{B}_r(x, y, e) \;\Leftrightarrow\; [A_r(x, y, e) + I_r(x, y, e)]/2 > thr$$

In words: a speaker categorizes an event e as an instance of a binary form with arguments x and y if the average level of activity and intentionality from x towards y in e exceeds the threshold *thr*. Our critical assumption is about the way binary ($\mathbf{B}_r$) predicates (1c/d), collective ($\mathbf{C}_r$) predicates (1a), and 'with' ($\mathbf{W}_r$) predicates (1b) differ while manifesting the CC of one and the same root. We propose:

$$\mathbf{C}_r(x + y, e) \;\Leftrightarrow\; [A_r(x, y, e) + I_r(x, y, e) + A_r(y, x, e) + I_r(y, x, e)]/4 > thr$$

Here the typicality of e as an instance of the collective form $\mathbf{C}_r$ is the average level of activity and intentionality in e from x towards y and from y towards x. Thus, the collective form differs from the binary form since both directions of A and I (from x to y and from y to x) are equally weighed. It also contrasts with our proposed typicality of the 'with' form $\mathbf{W}_r$, where the subject's activity and intentionality weigh more heavily than the 'with'-complement. However, unlike other objects, the 'with'-complement's activity and intentionality are assumed to have a non-zero weight c. Formally:

$$\mathbf{W}_r(x, y, e) \;\Leftrightarrow$$
$$(1-c) \cdot [A_r(x, y, e) + I_r(x, y, e)]/2 + c \cdot [A_r(y, x, e) + I_r(y, x, e)]/2 > thr,$$
$$\text{where } 0 < c < 0.5$$

Note that substituting $c = 0$ in this definition of the 'with' predicate gives the definition of the binary meaning $\mathbf{B}_r(x, y, e)$ above, whereas $c = 0.5$ yields the definition of the collective meaning $\mathbf{C}_r(x + y, e)$. Thus, these two meanings are viewed as limiting cases of the 'with' predicate meaning.

In the three different verbal forms we consider, our proposal lets the activity (A) and intentionality (I) features have different weights in relation to the entities x and y. However, we assume that the same conceptual root r defines the A and I attributes, and that the same threshold *thr* is relevant for categorizing events using the three forms. This predicts some standardly assumed entailments between sentences like (1a-d), but also some non-entailments, as stated and illustrated below:

E1. If the binary form symmetrically holds between x and y in some event e, that entails the 'with' form in both its directions. For example:

A fights B and B fights A (in e)

$\Rightarrow$ *A fights with B and B fights with A* (in e)

E2. If the 'with' form symmetrically holds between x and y in some event e, that entails the collective form in both its directions:

A fights with B and B fights with A (in e)

$\Rightarrow$ *A and B fight, and B and A fight* (in e)

NE1. For neither the collective form nor the 'with' form is it necessary that the binary form symmetrically holds between x and y:

A and B fight *or* *A fights with B* $\not\Rightarrow$ *A fights B and B fights A*

More generally, from our proposal it can be concluded that neither the collective nor the 'with' form entails that the binary form holds in any of its two directions:

(i) *A and B fight* $\not\Rightarrow$ *A fights B* (ii) *A and B fight* $\not\Rightarrow$ *B fights A*

(iii) *A fights with B* $\not\Rightarrow$ *A fights B* (iv) *A fights with B* $\not\Rightarrow$ *B fights A*

NE2. The collective form does not entail the 'with' form:

A and B fight $\not\Rightarrow$ *A fights with B*

NE3. The 'with' form does not entail the collective form:

A fights with B $\not\Rightarrow$ *A and B fight*

NE4. The 'with' form is not symmetric:

A fights with B $\not\Rightarrow$ *B fights with A*

NE5. The binary form does not entail the 'with' form:

A fights B $\not\Rightarrow$ *A fights with B*

The proofs of these (non-)entailments follow from our definitions of the $\mathbf{B}_r$, $\mathbf{C}_r$ and $\mathbf{W}_r$ predicates by simple numeric considerations.

Entailments *E1* and *E2* are consensual but express regularities that cannot be accounted for by merely giving an unanalysed theta role to the 'with' complement as in Rákosi's proposal [6]. The non-entailments *NE1-NE5* go against most previous proposals about lexical reciprocity. To examine these predictions, we collect more data on collective, binary and 'with' predicates using the same methods of [4].

3 The experimental work

3.1 Materials and procedure

Five Dutch verbs were examined: *vechten* 'fight', *knuffelen* 'hug', *praten* 'talk', *botsen* 'collide', and *appen* 'send WhatsApp'. All verbs have a *met* ('with') alternation and another binary alternation: transitive (e.g. *knuffelen*) or prepositional (e.g. *botsen tegen* 'collide against'). Kruitwagen et al. [4] report initial experiments with these verbs and video materials describing different asymmetric events. To evaluate the current, more general proposal, our experiments use the same video materials with a wider range of constructions, including the 'with' forms and more ordering variations between the arguments. Two videos clips were used per verb, featuring two human actors: 'Violet' and 'Mark'. In both clips Mark is passive while Violet is active and shows the expected intention. For example, in the video clips corresponding to

sentences with the verb *vechten* 'fight', Violet hits Mark angrily. In one of the two clips, which we refer to as 'Intentional Mark' (*M+I*), Mark shows the expected intention (e.g. anger). In the other clip, *M-I*, Mark looks indifferent. To the collective and binary sentences that were examined in [4] we added the other direction of the binary sentence and the two directions of the 'with' sentence (1b). This results in a total of three data points per verb and clip on top of the two examined by [4]. Thus, for each verb root we collected data in a 2×5 between-subject design with independent variables Intention of the passive figure 'Mark' (*M+I* or *M-I*), and Sentence Type of five different values:

V&M	collective form, active first	*Violet en Mark knuffelen* ('hug')
VbM	binary form, active subject	*Violet knuffelt* ('hugs') *Mark*
MbV	binary form, passive subject	*Mark knuffelt* ('hugs') *Violet*
VwM	'with' form, active subject	*Violet knuffelt met* ('hugs with') *Mark*
MwV	'with' form, passive subject	*Mark knuffelt met* ('hugs with') *Violet*

Each of 477 participants (286 female, age M=26) received a LimeSurvey questionnaire with four target items and six filler items, and made a truth-value judgement on those ten items based on the video clips.

		V&M		VbM		MbV		VwM		MwV	
		%	of	%	of	%	of	%	of	%	of
vechten	**M+I**	40%	53	77%	35	32%	57	62%	42	36%	39
('fight')	**M-I**	19%	58	55%	33	4%	49	39%	41	5%	40
knuffelen	**M+I**	84%	57	97%	32	40%	53	84%	45	61%	44
('hug')	**M-I**	51%	49	100%	33	9%	58	67%	43	37%	41
praten	**M+I**	35%	34	100%	32	0%	35	42%	43	16%	45
('talk')	**M-I**	16%	38	94%	33	0%	33	36%	44	0%	41
botsen	**M+I**	69%	59	100%	35	4%	53	94%	32	34%	32
('collide')	**M-I**	70%	53	91%	34	4%	53	82%	33	33%	33
appen	**M+I**	57%	35	100%	34	3%	33	74%	39	27%	41
('WhatsApp')	**M-I**	30%	33	100%	32	9%	34	24%	42	10%	40
TOTAL		**47%**	**469**	**91%**	**333**	**11%**	**458**	**60%**	**404**	**26%**	**396**

Figure 1: **Results** – *positive answers per clip and sentence form are reported as percentages ('%') of the total numbers of participants who responded to each truth-value judgement ('of').*

3.2 Results and Analysis

For each clip and sentence type per verb, Figure 1 reports the percentages of total number of participants who judged the sentence as true. Unsurprisingly, these results are consistent with the intuitive entailments *E1* and *E2*. More importantly, as can be seen in Figure 2, with 9 of the 10 scenarios (5 of 5 verbs), significantly more participants ($p < 0.05$, Fisher exact test 2-tailed) answered positively on the collective *V&M* sentences than on the binary *MbV* sentences (column (i)). Similar results appeared for the 'with' *MwV* sentences against the binary *MbV* sentences (6/10 scenarios, 4/5 verbs, column (ii)), and against the binary *VbM* sentences (10/10 scenarios, 5/5 verbs, column (iii)). This supports *NE1*. In a similar fashion, the results show support for *NE2* (6/10 scenarios, 4/5 verbs, column (iv)), *NE3* (4/10 scenarios, 3/5 verbs, column (v)), *NE4* (9/10 scenarios, 5/5 verbs, column (vi)), and *NE5* (5/10 scenarios,

		(i) V&M vs MbV	(ii) MwV vs MbV	(iii) MwV vs VbM	(iv) V&M vs MwV	(v) V&M vs VwM	(vi) VwM vs MwV	(vii) VwM vs VbM
vechten	**M+I**	p=0.428	p=0.666	**p<0.001**	p=0.829	**p=0.040**	**p=0.026**	p=0.217
('fight')	**M-I**	**p=0.034**	p=1	**p<0.001**	p=0.068	**p=0.039**	**p<0.001**	p=0.242
knuffelen	**M+I**	**p<0.001**	**p=0.042**	**p<0.001**	**p=0.012**	p=1	**p=0.018**	p=0.129
('hug')	**M-I**	**p<0.001**	**p<0.001**	**p<0.001**	p=0.204	p=0.139	**p=0.008**	**p<0.001**
praten	**M+I**	**p<0.001**	**p=0.016**	**p<0.001**	p=0.062	p=0.641	**p=0.009**	**p<0.001**
('talk')	**M-I**	**p=0.027**	p=1	**p<0.001**	**p=0.009**	**p=0.047**	**p<0.001**	**p<0.001**
botsen	**M+I**	**p<0.001**	**p<0.001**	**p<0.001**	**p=0.0018**	**p=0.008**	**p<0.001**	p=0.224
('collide')	**M-I**	**p<0.001**	**p<0.001**	**p<0.001**	**p=0.0016**	p=0.310	**p<0.001**	p=0.305
appen	**M+I**	**p<0.001**	**p=0.009**	**p<0.001**	**p=0.010**	p=0.144	**p<0.001**	**p=0.0012**
('WhatsApp')	**M-I**	**p=0.033**	p=1	**p<0.001**	**p=0.038**	p=0.603	p=0.142	**p<0.001**
TOTAL		**p < 0.00000001**	**p < 0.00001**	**p < 0.00001**	**p < 0.00001**	**p = 0.0001**	**p < 0.00001**	**p < 0.00001**

NE1 NE2 NE3 NE4 NE5

Figure 2: **Fisher exact tests (two-tailed)** – *the comparisons between responses on the five sentence forms evaluate the evidence for non-entailments* NE1-NE5.

3/5 verbs, column (vii)). The totals in Figure 2 show the same kind of support, and the trends in the tests of *NE1-NE5* are in the same direction for all verbs and scenarios.

3.3 Discussion

All the effects discovered support *NE1-NE5* and are consistent with *E1-E2*. We propose that this results from differences in argument structure as constrained by our model. Binary entries (e.g. transitive *knuffelen* 'hug') require higher levels of activity and intentionality of the singular agent compared to 'with' forms, which in turn require more agent activity and intentionality than what is required from each of the two agents in a conjunctive *A and B* subject of a collective intransitive form.

3.4 Followup

The non-entailments *NE1*(i-ii) involve a conjunctive subject (e.g. *Violet and Mark* VERB) with a collective verbal form, as well as a 'with' form with the opposite order (*Mark* VERB *with Violet*). This change in ordering might in principle provide an alternative account of *NE1*(i-ii): linear word order might have a general effect on agency requirements, both for two conjoined NPs and for two NPs with different grammatical functions. To examine this possible account we ran a similar procedure on Prolific comparing truth-value judgements on sentences of the forms *Violet and Mark*/*Mark and Violet vechten* ('fight'). With 32-40 (M=35) participants per condition, we did not discover any significant effect according to the Fisher Exact test, with maximal odd-ratio effect sizes for *vechten* 'fight' *M-I* (p=0.094, *o.r.*=2.92) and *knuffelen* 'hug' *M+I* (p=0.613, *o.r.*=3.1). Thus, we found no evidence that linear order alone explains the results regarding *NE1*(i-ii), though it might play a small role.

4 Conclusions

Our threshold model of reciprocal alternations provides a more nuanced analysis of theta-roles than what mere labels like 'agent', 'patient' or 'partner' describe. In this sense our proposal substantiates Dowty's [2] proto-role assumptions. Unlike previous proposals, our analysis is

supported by extensive experimental evidence. Future work may be able to extend this picture further by testing the separation between Activity and Intentionality and their linear contribution, e.g. by comparing our model to (possibly non-linear) typicality equations with other conceptual parameters.

References

[1] Alexis Dimitriadis. Irreducible symmetry in reciprocal constructions. In Ekkehard König and Volker Gast, editors, *Reciprocals and Reflexives: Theoretical and Typological Explorations*, pages 375–410. De Gruyter, Berlin, 2008.

[2] David Dowty. Thematic proto-roles and argument selection. *Language*, 67(3):547–619, 1991.

[3] James A. Hampton. Typicality, graded membership, and vagueness. *Cognitive Science*, 31(3):355–384, 2007.

[4] Imke Kruitwagen, Yoad Winter, and James Hampton. Reciprocal verbs: a prototype model. In Andrea Beltrama, Florian Schwarz, and Anna Papafragou, editors, *The first Experiments in Linguistic Meaning conference ELM1*, volume 1, pages 197–203. LSA, 2021.

[5] Beth Levin. *English verb classes and alternations: A preliminary investigation.* University of Chicago Press, Chicago, 1993.

[6] György Rákosi. The inherently reflexive and the inherently reciprocal predicate in Hungarian: Each to their own argument structure. In Ekkehard König and Volker Gast, editors, *Reciprocals and Reflexives: Theoretical and Typological Explorations*, pages 411–450. De Gruyter, Berlin, 2008.

[7] Malka Rappaport-Hovav and Beth Levin. Building verb meanings. In Miriam Butt and Wilhelm Geuder, editors, *The projection of arguments: Lexical and compositional factors*, pages 97–134. CSLI Publications, Stanford, CA, 1998.

[8] Tal Siloni. Reciprocal verbs and symmetry. *Natural Language & Linguistic Theory*, 30(1):261–320, 2012.

Aboutness and Modality*

Dean McHugh

ILLC and Department of Philosophy, University of Amsterdam

In this paper I would like to offer a new framework for hypothetical reasoning, with the goal of predicting what scenarios we consider when we interpret a conditional or causal claim (such as a sentence containing the word *because*). The idea is that when we interpret a conditional or causal claim, we identify a part of the world to change and imagine changing that. Sentences are about parts of the world: when we interpret a conditional antecedent or *because* clause, we allow the part of the world it is about to vary. This expands our modal horizons, which we restrict to those scenarios where the sentence we are asked to imagine true is indeed true. To evaluate a whole conditional or causal claim we look to the possible futures after this change.

My main evidence for this approach is that it gives just the right range of hypothetical scenarios to account for how we interpret both conditionals and causal claims. Some approaches (such as Stalnaker and Lewis's semantics of conditionals based on similarity, and Kratzer's premise semantics) consider too few scenarios, while others (such as Fine's 2012 truthmaker semantics of conditionals) consider too many. The present approach inhabits a Goldilocks zone between these extremes: not too restrictive, not too permissive, but just right.

Section 1 motivates giving a uniform account of modality for conditionals and causal claims. In Sections 2 and 3 we address a challenge for uniformity. Section 4 uses aboutness to model how sentences raise hypothetical scenarios and compares this approach with alternatives.

1 Modal uniformity

A compelling idea in the analysis of causality is that conditionals and causal claims raise hypothetical scenarios in a uniform way; in other words, the two constructions activate the same general faculty of hypothetical reasoning. Call this idea *modal uniformity*. One piece of evidence for modal uniformity is that we so often paraphrase one construction with the other. David Hume inaugurated counterfactual approaches to causation in 1748 when he wrote

> We may define a cause to be an object followed by another ... where, if the first object had not been, the second never had existed. (*Enquiry*, Section VII)

The parallel between conditionals and causal claims lives on today in the panoply of counterfactual analyses of causation.[1] Here is a more contemporary illustration of the close connection between conditionals and causal claims. Title VII of the U.S. Civil Rights Act states

> It shall be an unlawful employment practice for an employer to fail or refuse to hire or to discharge any individual ... because of such individual's race, color, religion, sex, or national origin. (78 Statute 241, Section 703(a)(1), p. 255)

The text uses a causal word: *because*. Now here is Justice Elena Kagan discussing the same law at oral argument for a 2019 U.S. Supreme Court case, *Bostock v Clayton County*.

> Kagan: What you do when you look to see whether there is [sex] discrimination under Title VII is, you say, would the same thing have happened to you if you were of a different sex? (Oral argument, pp. 41–42)

*Thanks to Katrin Schulz, Lorenzo Pinton and Robin Martinot for valuable comments. All errors are mine.

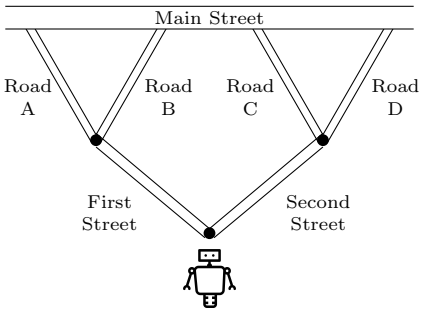
[1] Examples include Lewis (1973a, 2000), the approaches in Collins, Hall, and Paul (2004), and many more.

Notice how Kagan uses a conditional ("if you were of a different sex") to express a causal claim ("because of ... sex"). A causal claim winds up as a conditional. In fact, Title VII does not contain any conditionals of the kind uttered by Kagan. If conditionals and causal claims did not raise hypothetical scenarios in a uniform way, the fact that we so often paraphrase one construction with the other—as David Hume and Elena Kagan did—would be mysterious.

There is one main challenge to modal uniformity: sufficiency, which we turn to now.

2 Sufficiency

An often overlooked aspect of the meaning of *cause* and *because* is that they imply that the cause was in some sense sufficient for the effect. To illustrate, imagine a robot has to get to Main Street (see right). At each fork it turns at random. On this particular day, it took First Street and then Road B. Consider (1) in this context.



(1) a. The fact that the robot took First Street caused it to take Road B.
b. The robot took Road B because it took First Street.

(1) are unacceptable in this context. Intuitively this is because after taking First Street, the robot could have taken Road A instead. This suggests the following, informal principle.

Sufficiency requirement. *E because C* and *C cause E* entail that *C* is sufficient for *E*.

To test this, let us minimally change the context to make taking First Street sufficient for taking Road B. Imagine, say, that the robot has been programmed to always change direction. For example, if it turns left at one fork it must turn right at the next. Consider (1) in this context.

(1) are much better. Counterfactual dependence does not account for the contrast between the turn-at-random and always-change-direction contexts. For in both, if the robot hadn't taken First Street it wouldn't have taken Road B. The difference is rather a difference in sufficiency.

We can account for this contrast by incorporating the openness of the future into the analysis of sufficiency. Informally, let us say that for sentence *A* to be sufficient for sentence *C*, *C* must be true in every nomically possible future after *A* becomes true.[2] When the robot turns at random, but not when it is programmed to always change direction, it is nomically possible for it to take Road A after First Street. If we then assume that *C cause E* and *E because C* imply that *C* is sufficient for *E*, we correctly predict (1) to be unacceptable when the robot turns at random but acceptable when it is programmed to always change direction.[3]

Some violations of sufficiency do not involve the openness of the future. Consider (2)–(4).

(2) *Yves only buys a specific shade of blue paint: ultramarine. In a paint shop he sees paints in various colours, including various shades of blue. He bought some ultramarine paint.*
a. Yves bought this paint because it is blue.
b. The fact that this paint is blue caused Yves to buy it.

(3) *Alice is 16 and tried to get into a bar for over 18s. The bouncer did not let her in.*

[2]Semantics of conditionals that use the openness of the future include Thomason (1970), Thomason and Gupta (1980), Arregui (2005), Ippolito (2013), Khoo (2015), and Canavotto (2020).

[3]We assume we have some way to extract sentence *C* from the noun and sentence *E* from the *to*-infinitive.

a. The bouncer did not let Alice in because she is under 30.
b. The fact that Alice is under 30 caused the bouncer to not let her in.

(4) *Ali was born in Ireland and has Irish citizenship.*
a. Ali has an Irish passport because he was born in Europe.
b. The fact that Ali was born in Europe caused him to get Irish citizenship.

These sentences greatly improve when we replace the cause with a minimally different one that is sufficient for the effect; e.g. replacing *blue* with *ultramarine*, *under 30* with *under 18*, and *Europe* with *Ireland*. These contrasts are not due to a difference in counterfactual dependence. For example, if Alice had been over 18, or over 30, she would have gotten in. Nor are they due to the openness of the future. Accounting for, say, (2)'s unacceptability in this way would require a time when the paint was blue but not yet determined that it would be ultramarine. But specific shades do not come after general shades in time – every blue is a kind of blue.

Given modal uniformity, a natural idea is that A is sufficient for C just in case the *would*-conditional "if A, would C" (denoted $A > C$) is true. A problem is that almost all semantics of *would*-conditionals validate conjunctive sufficiency, that $A \wedge C$ entails $A > C$.[4] This is a natural principle on similarity approaches to conditionals. It follows from *strong centering*: the assumption that no world is as similar to any world w as w is to itself. But then we cannot say that A is sufficient for C just in case $A > C$ is true. For example, when the robot turned at random, it took First Street and Road B, but taking First Street was not sufficient for it take Road B. Even though Ali was born in Europe and has an Irish passport, being born in Europe is not sufficient to have an Irish passport.

3 Selection functions to the rescue

There is, however, a way out. For there is no evidence that the strong centering requirement of conditionals must come from our general capacity to imagine hypothetical scenarios (say, from the nature of similarity). Indeed, Mandelkern (2018) and Cariani and Santorio (2018) and Cariani (2021) propose that strong centering has another source: the meaning of *will* and *would* involves a selection function which chooses a world from a set of worlds. The only additional constraint is that if the selection function can choose the world of evaluation, it must.

(5) A selection function $s : (W \times \wp(W)) \to W$ takes a world w and a set of worlds p and returns a world, where $s(w, p) \in p$ (*success*) and if $w \in p$ then $s(w, p) = w$ (*centering*).

Selection functions allow us to rescue modal uniformity. For we can propose that we have a general capacity to consider hypothetical scenarios in response to a sentence – one we use to interpret both conditionals and causal claims – that is not strongly centered. In addition, *would* contributes a selection function but *cause* and *because* do not. Put another way, *would*-conditionals and *(be)cause* have the same modal horizon but a different modal force.

Some evidence that *would* and *(be)cause* have a different modal force comes from bets and probability judgements. Suppose the robot took First Street. Consider (6).

(6) a. Ali: "If the robot had taken Second Street, it would have taken Road C because it took Second Street."
b. Bob: "If the robot had taken Second Street, it taking Second Street would have caused it to take Road C."

[4] E.g. Stalnaker (1968), Lewis (1973b), Pollock (1976), and Kratzer (2012) validate conjunctive sufficiency.

c. Carl: "If the robot had taken Second Street, it would have taken Road C."

Suppose Alison says, "I bet that what Ali said is true." Bea says: "I bet that what Bob said is true" and Cleo says, "I bet that what Carl said is true." The three are then told that the robot was programmed to turn at random. Who won their bet? Intuitively, Alison and Bea lost. This is expected if *cause* and *because* quantify universally over possible futures after the cause occurs. In contrast, Cleo's bet intuitively remains undecided. A number of authors observe that this is unexpected if Carl's utterance meant that in *all* possible futures after the robot takes Second Street, it takes Road C – since this is clearly false.[5] Similar remarks apply concerning probability. Consider: what is the probability that what Ali/Bob/Carl said is true?[6] Intuitively, Ali and Bob's sentences have a much lower probability of being true than Carl's.

In light of these data, we propose that sufficiency involves universal quantification over a set of worlds, while *would*-conditionals select a world from this set. Schematically, for any world w and sentence A, we let $mh(w, A)$ be this set of worlds, which following von Fintel (2001) we call the *modal horizon*. For any sentence A let $|A|$ be the set of worlds where A is true. We propose the following clauses for sufficiency ($\gg$) and *would*-conditionals ($>$).

(7) Where mh is the modal horizon function and s the selection function,

$$\begin{array}{lll} A \gg C \text{ is true at } w & \text{if and only if} & mh(w, A) \cap |A| \subseteq |C| \\ A > C \text{ is true at } w & \text{if and only if} & s\big(w, mh(w, A) \cap |A|\big) \in |C| \end{array}$$

These clauses allow us to have our cake and eat it. They preserve modal uniformity while only selectively validating conjunctive sufficiency: $A \wedge C$ entails $A > C$ but not $A \gg C$ (assuming every world is in its own modal horizon, i.e. $w \in mh(w, A)$, a fact we prove below).

How does sufficiency fit into the overall meaning of *cause* and *because*? McHugh (forthcoming) analyses their meaning in terms of two relations between sentences, sufficiency and production, whereby E *because* C and C *cause* E are equivalent to the following.

(8) **Meaning of *(be)cause*:** $C \wedge \big(C \gg (C \textit{ produce } E)\big) \wedge \neg\big(\neg C \gg (\neg C \textit{ produce } E)\big)$.

Here we will only discuss the sufficiency requirement. It follows from the fact that $\gg$ validates right weakening: if A is sufficient for C and C entails D then A is sufficient for D. Then assuming production is factive (C *produce* E entails $C \wedge E$), we derive the sufficiency requirement:

$$E \textit{ because } C \quad \Rightarrow \quad C \gg (C \textit{ produce } E) \quad \Rightarrow \quad C \gg (C \wedge E) \quad \Rightarrow \quad C \gg E.$$

4 The modal horizon

Our question now is how to analyse the modal horizon. In other words, what hypothetical scenarios do we consider when we evaluate a conditional or causal claim? To give a flavour of the approach I would like to adopt, consider the image on the right, of an orange circle on a blue background. Imagine the circle with a different colour. What images come to mind?



[5] This observation regarding conditionals in betting contexts has been discussed by Prior (1976:100), Moss (2013), Belnap, Perloff, and Xu (2001:160), Cariani and Santorio (2018) and Cariani (2021:63).

[6] This way of formulating the question follows a strategy by Mandelkern (2018) to avoid the objection that conditionals are not assigned probabilities. When one is asked "What is the probability that what x said is true?" where x said a conditional, one is asked to assign a probability to an entire conditional, rather than a probability to the consequent while supposing the antecedent. For discussion see Mandelkern (2018:304–305).

Here are some images we may consider (left), and some images we do not consider (right).



Remarkably, we all seem to imagine the same kinds of images. We are only given the image and the sentence "imagine the circle with a different colour". We are not explicitly told what stays the same, but figure that out ourselves. How do we do this?

One idea – inspired by similarity approaches to conditionals – is that we rank all possible images in terms of similarity to the actual image. Another – inspired by premise semantics – is that we identify a set of true propositions, and seek to preserve the truth of as many of these as possible while maintaining consistency with the circle being different colour.

On second thought, however, when we are asked to imagine the circle a different colour, intuitively we do not rank all images by similarity, nor do we compare all images in terms of what propositions they make true. Rather, we identify a part of the image that needs to change and change that. All else is background, so to speak, so it is kept the same – it is the *ceteris* in *ceteris paribus*.[7] Let us therefore start with a simple question: when we imagine the circle a different colour, what parts of the image change, and what parts stay the same?

We can answer this question without much thought at all. Let us list some parts of the image and check for each whether it stays the same when we imagine the circle a different colour. This is illustrated in Figure 1. For example, the top half of the image does not stay the same. Looking at the Figure, we see that a part of the image stays the same just in case it does not overlap the circle (where two parts overlap just in case they have a part in common).

Let us assume we have a partial order $(S, \leq)$ where the elements of S represent parts of the world at a moment in time, and $\leq$ represents parthood. We assume that every state is part of a maximal state (in Fine's 2017 terminology, our state spaces are *world-spaces*). We further assume that we have an *aboutness* relation between states and sentences.[8] For example, *The circle is a different colour* (A) is only about the circle, and not, say, about the top half of the image (our aboutness relation represents 'whole' aboutness). We define that a state s is in the *background* of sentence A just in case s does not overlap any state A is about. For any sentence A and maximal states t and t', let us call t' an A-*variant* of t just in case every part of t in the background of A is part of t'. This is illustrated in Figure 2. Informally, t's A-variants result from removing the parts of t that A is about; they agree with t on everything A is not about. Put another way, *ceteris paribus* means having the *ceteris* – the background states – as parts.

To illustrate, images I–III are A-variants of the original image, but IV and V are not. And A is false at III. So restricting to the A-variants of the original image where the circle is indeed a different colour allows I and II but rules out III–V, as desired. This is depicted below, where the grey squares represent the absence of a state (rather than the presence of grey squares).



While this example involves parthood in space, our analysis also applies to parthood broadly construed. For example, when we imagine a red ball blue, we intuitively keep its shape. We

[7] A 'foreground'/'background' distinction has also been used by Ciardelli, Zhang, and Champollion (2018).
[8] Some recent work on aboutness is Yablo (2014), Fine (2017b), Berto (2018), Hawke (2018) and much more.

can predict this on the present approach if an object's shape and colour are non-overlapping parts of it. Given that the *The ball is blue* is not about its shape, its shape is in this sentence's background and therefore part of every variant of the ball we consider when we imagine it blue.

We also reason hypothetically in a changing world. Suppose a light switch is up and the light connected to it is off. When we interpret *The light is off because the switch is up*, or *If the switch were down, the light would turn on*, the *because* clause and antecedent are only about the switch, but we imagine the light changing too. Intuitively, the sentence $A =$ *The switch is up/down* is about the state of the switch. This switch does not overlap the state of the light, mereologically speaking; see Figure 3(a). So the state of the light being off is in A's background, and is therefore part of the A-variants of the current state. To reach a scenario where the light turns on, we need to specify how things change through time. The simplest way to represent time, I believe, is via linear orders. The maximal states of our state space, called *moments*, represent how things stand at a point in time. We take a world to be a linear order of moments. Finally, our models specify which worlds are nomically possible and which are not. Figure 3(b) illustrates the nomic possibilities for the light switch. In general, we define:

Definition 1 (Nomic aboutness model)**.** *Where S is a set and $\leq$ a binary relation on S, define:*

$$\begin{aligned} &\mathit{Sit} := S \times I, \text{ where } I \text{ is an arbitrary label set},\\ &M := \{t_i \in \mathit{Sit} : t \leq u \text{ implies } t = u \text{ for all } u \in S\},\\ &W := \{(M', \preceq) : M' \subseteq M,\ \preceq \text{ is a linear order}\}. \end{aligned}$$

Given a set of sentences $\mathcal{L}$, a nomic aboutness model is a tuple $(S, \leq, \mathcal{A}, P, |\cdot|)$ where $(S, \leq)$ is a partial order such that every state is part of a moment, $\mathcal{A} \subseteq \mathcal{L} \times S$, $P \subseteq W$, and $|\cdot| : \mathcal{L} \to W$.

Sit represents the set of situations, where a situation is a particular instance of a state. (We need situations since the same state may repeat at a world.) M represents the set of moments (maximal situations) and W the set of logically possible worlds. $\mathcal{A}(A, s)$ says that sentence A is about state s, $w \in P$ that world w is nomically possible, and $w \in |A|$ that A is true at w. In these terms, for any moments t, t' and sentence A, t' is an A-variant of t just in case $\forall s \leq t\big((\forall u \in S : \mathcal{A}(A, u) \Rightarrow s \not\circ u) \Rightarrow s \leq t'\big)$, where $s \not\circ u$ denotes that s and u do not overlap.

Following recent work in tense-modal interaction (e.g. Condoravdi 2002), we assume a historical modal base: all worlds in $mh(w, A)$ share w's past. We also assume that we access hypothetical alternatives a point in time, what we call *intervention time*.[9] We propose that the modal horizon contains the possible futures of each A-variant at the intervention time, while leaving the past unchanged. Formally, for any world w and moment t on w, let $w_{\prec t}$ be the segment of w up to but not including t, let $w_{\succeq t}$ be the sequence of w from t on, including t, and let $\frown$ denote concatenation.[10] We define the modal horizon as follows, illustrated in Figure 4.

Definition 2 (The modal horizon)**.** *For any sentence A, moment $t \in M$ and world $w \in W$,*

$$mh_t(w, A) := \{w_{\prec t} \frown w'_{\succeq t'} : t' \text{ is an } A\text{-variant of } t, t' \in w' \text{ and } w' \in P\}.$$

Informally, the modal horizon results from (i) making a sudden change by allowing everything A is about at intervention time to vary – determined by the parthood and aboutness relations – (ii) taking the nomically possible futures after this change, and (iii) gluing on the actual past.

[9] Condoravdi (2002) calls this the *temporal perspective*, Arregui (2005) and Ippolito (2013) call it the *accessibility time*, Khoo (2015) calls it the *modal time* and Khoo (2017) calls it the *counterfactual time*.

[10] Concatenation is defined by: $(X, \preceq) \frown (X', \preceq') = (X \uplus Y, \preceq \frown \preceq')$, where $X \uplus Y$ is the disjoint union of X and Y, and $x \preceq \frown \preceq' y$ just in case (i) $x, y \in X$ and $x \preceq y$, or (ii) $x, y \in Y$ and $x \preceq' y$, or (iii) $x \in X$ and $y \in Y$.

Appendix: Comparison with alternative frameworks

Similarity approaches are too restrictive

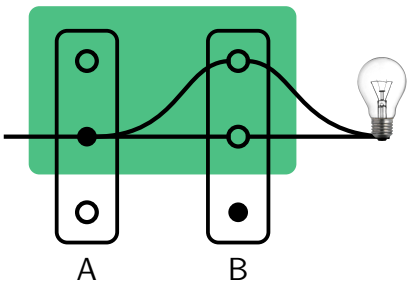
As we saw in Section 2, given strong centering (i.e. that $w <_w w'$ for all distinct worlds w, w'), if we took the modal horizon $mh(w, A)$ to consist of all the most similar worlds to w where A is true, we would undesirably predict that whenever A and C are true, A is sufficient for C.

In contrast, our construction of the modal horizon blocks this entailment, and can therefore account for the sufficiency violations in (1)–(4), both those due to the openness of the future and those due to the various ways for sentences to be true. For example, where *First* = *the robot took First Street*, and t is any moment before the robot took First Street, the only *First*-variant of t is t itself (that is, there is only one way for the robot to take First Street), but there is a nomically possible future after t where it does not take Road B. We therefore predict taking First Street was not sufficient for the robot to take Road B, as desired. Turning to (2), for example, we assume *Blue* = *The paint is blue* is about the state of the paint's colour (more specifically in the case, the state representing that the paint is blue). It is natural to suppose that this state is the same as – or at least overlaps – the state of the paint being ultramarine. Thus the *Blue*-variants allow the state of the paint being ultramarine to vary. We then restrict to those worlds containing the *Blue*-variants where the paint is indeed blue. We therefore correctly account for the fact that we consider various shades of blue when we interpret (2).

Nonetheless, on our approach $A \wedge C$ entails $A > C$ since every world is in its own modal horizon: $w \in mh_t(w, A)$ for every moment t, world w and sentence A. This holds since every moment is an A-variant of itself, every world is the concatenation of its past, present and future ($w = w_{\prec t} \frown w_{\succeq t}$), and the world of evaluation is itself nomically possible ($w \in P$). Then given the selection function's centering requirement, if A is true at w then $s\big(w, mh_t(w, A) \cap |A|\big) = w$.

There is independent evidence from *would*-conditionals that the modal horizon given by similarity approaches is too narrow. Consider the setup below: there are two switches, A and B, connected to a light. Part of the image is shaded. Each switch has three possible positions: up, in the middle, or down. As the wiring indicates, the light is on just in case A is in the middle and B is either up or in the middle. Currently, A is in the middle and B is down, so the light is off. Consider (9) in this context.

(9) a. If switch B were in the shaded region, the light would be on.
b. If switch B were in the shaded region, both switches would be in the shaded region.
c. If both switches were in the shaded region, the light would be on.



(9a) and (9b) are clearly acceptable, but (9c) is dubious. Its interpretation is subtle and would certainly benefit from empirical testing. Nonetheless, one does not want to say that (9c) follows from (9a) and (9b) as a matter of logic. However, given the equivalence of (*both* $A \wedge B$) $\wedge B$ and (*both* $A \wedge B$), (9) is an instance of *cautious monotonicity*: the inference from $A > B$ and $A > C$ to $(A \wedge B) > C$. On similarity approaches (e.g. Lewis 1973b) this follows from reflexivity and transitivity of similarity, the bare minimum constraints to impose, without which the approach would be deeply unworkable. In contrast, the present approach can predict (9a) and (9b) to be true while (9c) is false. We simply say *switch B is in the shaded region* is about the state of switch B, while *both switches are in the shaded region* is about the state of both switches.

Kratzer's (2012) premise semantics is too restrictive

Kratzer's semantics of conditionals does not provide the range of scenarios we need to account for the sufficiency violations in (2)–(4). Kratzer (2012:132–33) works in a situation semantics, where situations are parts of possible worlds and each proposition is taken to be the set of situations where it is true. Kratzer (1989, 2012) defines that a proposition p *lumps* a proposition q at a world w just in case p is true at w, and every situation that is part of w where p is true, q is true. Kratzer then proposes the following semantics of *would*-conditionals. We begin with a set F_w of true propositions, called the *base set*, and a proposition p, expressed by the conditional antecedent. Let $F_{w,p}$ be the set of all consistent subsets of $F_w \cup \{p\}$ that (i) contain p and (ii) are closed under lumping. A *would*-conditional with antecedent p and consequent q is true at world w just in case every set in $F_{w,p}$ has a superset in $F_{w,p}$ that logically implies q.

Given modal uniformity, we would like this proposal to predict how causal claims raise hypothetical scenarios. Take (2). Let b and u be the propositions that the paint is blue and ultramarine, respectively. In a world w where the paint is ultramarine, b lumps u: the paint is blue in w, and every situation part of w where the paint is blue is a situation where it is ultramarine. Then for any base set F_w and subset A of $F_{w,b}$ whatsoever, by (i), A contains b, so by (ii), A contains u. So every element of $F_{w,b}$ entails that the paint is ultramarine. Kratzer's approach ignores the diverse shades of blue that we intuitively consider when we interpret (2).

Kratzer's premise semantics also validates cautious monotonicity, and therefore faces the problem in (9), predicting that (9a) and (9b) entail (9c). Recall that cautious monotonicity follows from reflexivity and transitivity of the similarity order. If we construct the order as Kratzer (1981:47) proposes, whereby $v \leq_w v'$ holds just in case every proposition in the ordering source at w that is true at v' is also true at v, then reflexivity and transitivity of the order follow, respectively, from reflexivity and transitivity of implication – bedrock principles of logic.

Fine's (2012) truthmaker semantics of conditionals is too permissive

Fine (2012) proposes a semantics of counterfactuals that takes into account the various ways for the antecedent to be true. His account has three primitives: a state space (i.e. states ordered by parthood), an exact verification relation between states and sentences, and a world-relative possible outcome relation between states. His clause is that $A > C$ is true at a world w just in case for every exact verifier t of A and possible outcome u of t at w, u contains an exact verifier of C. This semantics satisfies a principle Fine calls *Universal Realisability of the Antecedent*: if a counterfactual is true then it is true for any way in which its antecedent is true.

This semantics appears to be too permissive. Fine (2017c) proposes, quite plausibly, that every exact verifier of A is an exact verifier of $A \vee B$, and analogously, that every exact verifier of $A(c)$ is an exact verifier of $\exists x Ax$. Paired with Fine's semantics this validates simplification, the inference from $(A \vee B) > C$ to $(A > C) \wedge (B \wedge C)$, as well as the inference from $(\exists x Ax) > C$ to $\forall x(Ax > C)$. As is well-known, there are data challenging these inferences, such as the classic (10) adapted from McKay and Inwagen (1977), and (11) from Embry (2014).

(10) a. If Spain had joined the Allies or the Axis, they would have joined the Axis.
b. If Spain had joined the Allies, they would have joined the Axis.

(11) a. If it had snowed yesterday, I would have gone skiing.
b. If it had snowed 100 feet yesterday, I would have gone skiing.

Fine's semantics undesirably predicts that (10a) entails (10b). The problem is not unique to disjunction. In event semantics it is standardly assumed that declarative sentences existentially

quantify over events (Davidson 1967, who traces the idea to Reichenbach 1947). For example, *It snowed yesterday* is interpreted as $\exists e(e \textit{ is a snowing event} \wedge \textit{runtime}(e) \subseteq \textit{yesterday})$. Since every snowing-100-feet event is a snowing event, Fine also predicts that (11a) entails (11b).

A prominent reply on behalf of simplification is that when we judge (10) true, we assume that Spain joining the Axis is in some sense not genuinely possible (Warmbrōd 1981, Fine 2012, Starr 2014, Willer 2018). However, Lassiter (2018) observes that simplification can fail even when we regard both disjuncts as possible. Consider (12), based on an example of Lassiter's.

(12) Ali: If Spain had joined the Allies or the Axis, they would have joined the Axis.
Bob: I think that's likely, but not certain.

Bob clearly considers both disjuncts possible. But given that entailment preserves likelihood (see Yalcin 2010:921), if simplification were valid his utterance would imply (10b) is also likely.

A second problem with this response comes from the (13), due to Nute (1975).

(13) If we had had good weather last summer or the sun had grown cold,
we would have had a bumper crop.

This sentence is unacceptable, intuitively because when we interpret it we consider scenarios where the sun grows cold, in which case we do not have a bumper crop. Now, if (10) is fine because we take Spain joining the Axis to be a counterfactual impossibility, the challenge is to explain why we cannot rescue (13) in the same way. Surely the sun growing cold last summer is far more of an impossibility than WWII Spain joining the Allies.

The present proposal does not face this problem, since it does not validate simplification. Consider (10a). What, intuitively, does it say? This is reasonably clear: it says that WWII Spain were disposed to join the Axis over the Allies. Let us include this preference in our model: our state space contains a state, AXIS PREFERENCE, such that in every nomically possible world containing this state where Spain fights, they join the Axis (see Figure 5 for such a state space). Simplificaton fails assuming that *Spain joins the Allies* is about AXIS PREFERENCE but *Spain joins the Allies or the Axis* is not.[11] For then we keep AXIS PREFERENCE when we interpret (10a) but not (10b), which allows (10a) to be true while (10b) is false.

It remains to say why (10a) is acceptable but (13) is not. We propose that the key difference between them is that, while WWII Spain was disposed to join the Axis over the Allies, there is no state disposing the world to have good weather last summer over the sun growing cold. It is easy to imagine WWII Spain harbouring Axis sympathy, but bizarre to imagine a part of the world manifesting a preference between good weather and the sun growing cold last summer. Since there is no such preference in our state space, our account of (10a)'s acceptability does not inadvertently extend to predict that (13) is also acceptable.

[11] It is natural to wonder whether we can find general principles ensuring that *Spain joins the Allies* is about AXIS PREFERENCE, but *Spain joins the Allies or the Axis* is not. It turns out there are. Let us say that state s is *nomically relevant* to sentence A just in case: A is true at every nomically possible world containing s, or A is false at every nomically possible world containing s, or A is true at every nomically possible world not containing s, or A is false at every nomically possible world not containing s. (This definition is inspired by the definition of orthogonality from Lewis 1988. Formally, world $w = (M', \preceq)$ *contains* state s iff $s \leq t$ for some $t_i \in M'$.) We propose that nomic relevance is necessary for aboutness, and minimal nomic relevance is sufficient:

(i) a. If s is minimally nomically relevant to A then A is about s.
b. If A is about s then s is nomically relevant to A.

AXIS PREFERENCE is nomically relevant to *Spain joins the Allies* – since every nomically possible world containing the state is one where the sentence is false – and assuming the state is atomic, it is minimally so. But AXIS PREFERENCE is not nomically relevant to *Spain joins the Allies or the Axis*. The principles in (i) therefore ensure that *Spain joins the Allies* is about AXIS PREFERENCE but *Spain joins the Allies or the Axis* is not.

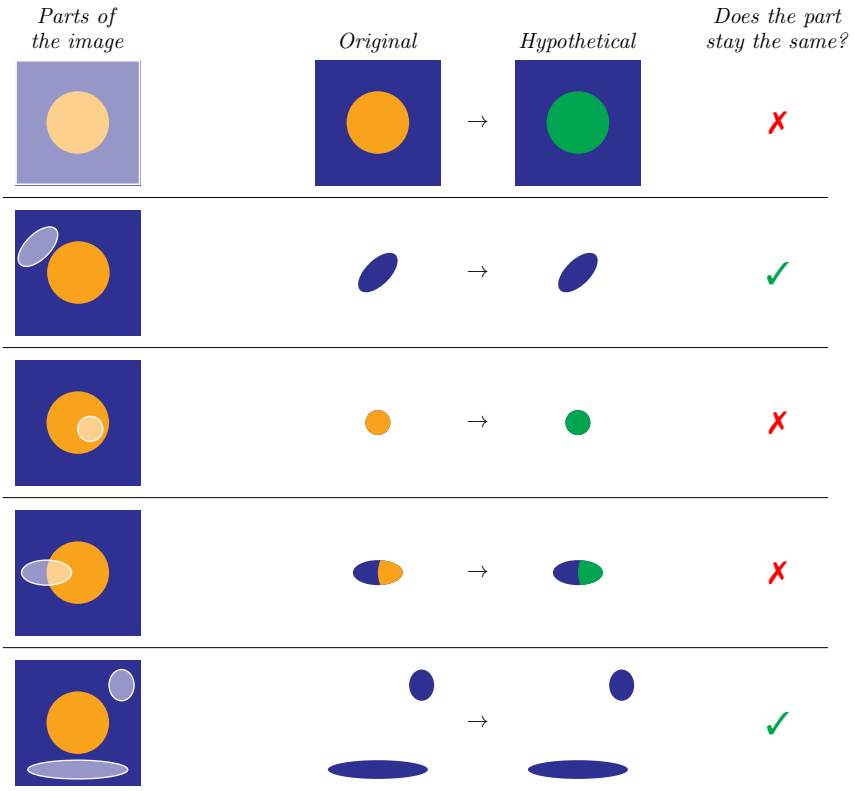


Figure 1: A part of the image stays the same just in case it does not overlap the circle.

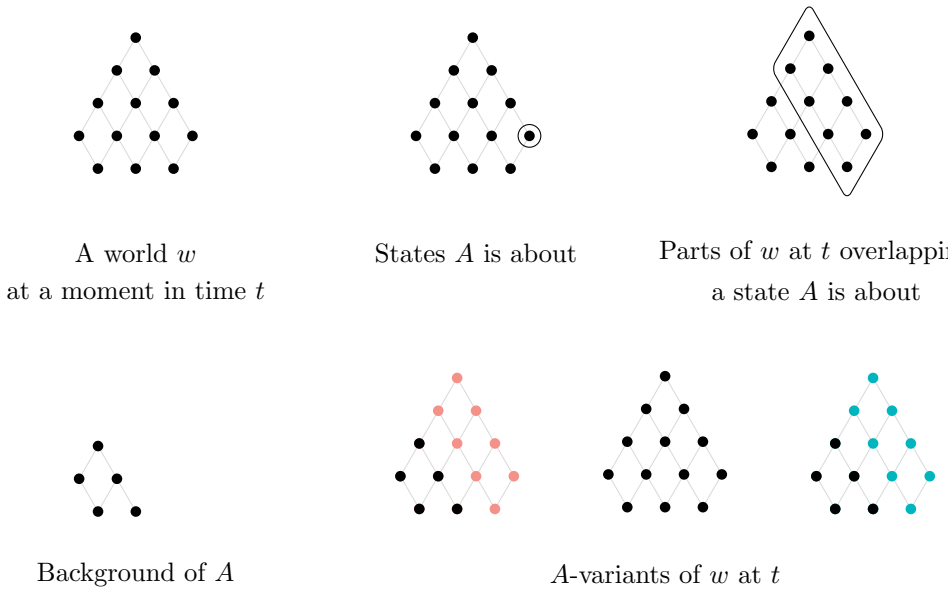
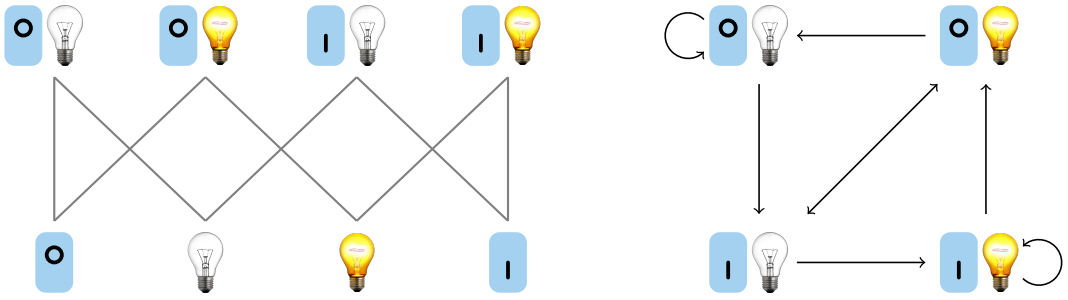


Figure 2: Steps to construct the A-variants of a world at a moment in time



(a) Mereological structure. (b) Nomic possibilities.

Figure 3: Light switch example. Nomically possible worlds correspond to directed paths in (b).

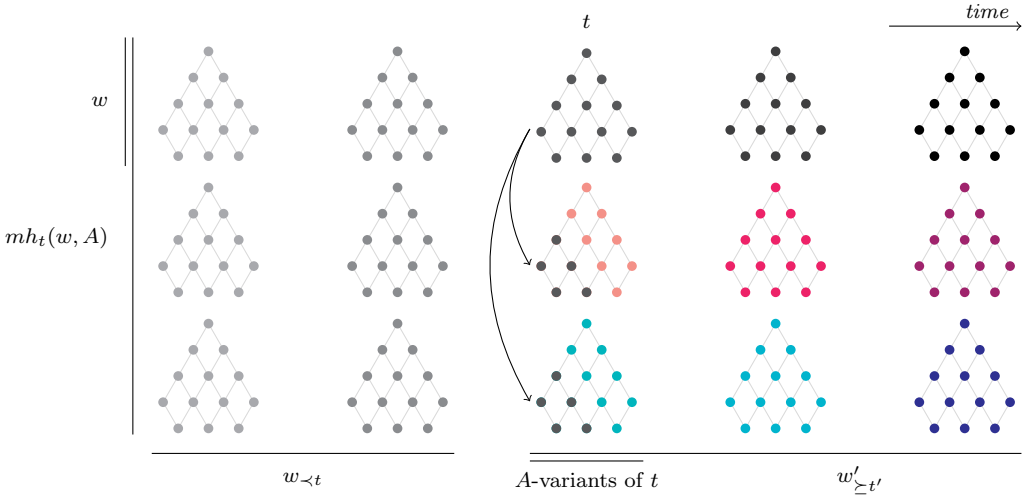


Figure 4: Constructing the modal horizon.

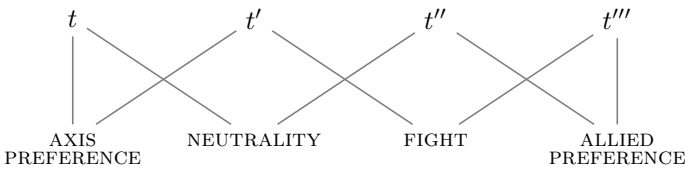


Figure 5: A toy state space for McKay and Inwagen's example. *Spain joins the Allies or Axis* is about NEUTRALITY, while *Spain joins the Axis* is about AXIS PREFERENCE and NEUTRALITY.

References

Arregui, Ana (2005). On the accessibility of possible worlds: The role of tense and aspect. PhD thesis. University of Massachusetts Amherst. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.7358&rep=rep1&type=pdf.

Belnap, Nuel, Michael Perloff, and Ming Xu (2001). *Facing the future: agents and choices in our indeterminist world*. Oxford University Press.

Berto, Francesco (2018). Aboutness in imagination. *Philosophical Studies* 175.8, pp. 1871–1886. DOI: 10.1007/s11098-017-0937-y.

Canavotto, Ilaria (2020). Where responsibility takes you: Logics of agency, counterfactuals and norms. PhD thesis. University of Amsterdam. DOI: 10.1007/978-3-031-17111-6.

Cariani, Fabrizio (2021). *The Modal Future: A Theory of Future-Directed Thought and Talk*. Cambridge University Press.

Cariani, Fabrizio and Paolo Santorio (2018). *Will* done better: Selection semantics, future credence, and indeterminacy. *Mind* 127.505, pp. 129–165. DOI: 10.1093/mind/fzw004.

Ciardelli, Ivano, Linmin Zhang, and Lucas Champollion (2018). Two switches in the theory of counterfactuals. *Linguistics and Philosophy* 41.6, pp. 577–621. DOI: 10.1007/s10988-018-9232-4.

Collins, John, Ned Hall, and Laurie A Paul (2004). Counterfactuals and Causation: History, problems, and prospects. *Causation and Counterfactuals*. MIT Press.

Condoravdi, Cleo (2002). Temporal Interpretation of Modals: Modals for the Present and for the Past. *The Construction of Meaning*. CSLI Publications, pp. 59–88. URL: https://semanticsarchive.net/Archive/2JmZTIwO/temp-modals.ps.

Davidson, Donald (1967). The logical form of action sentences. *The logic of decision and action*. Ed. by N. Rescher. University of Pittsburgh Press.

Embry, Brian (2014). Counterfactuals without possible worlds? A Difficulty for Fine's exact semantics for counterfactuals. *The Journal of Philosophy* 111.5, pp. 276–287. DOI: 10.5840/jphil2014111522.

Fine, Kit (2012). Counterfactuals without possible worlds. *Journal of Philosophy* 109.3, pp. 221–246. DOI: 10.5840/jphil201210938.

— (2017a). A theory of truthmaker content I: Conjunction, disjunction and negation. *Journal of Philosophical Logic* 46.6, pp. 625–674. DOI: 10.1007/s10992-016-9413-y.

— (2017b). A theory of truthmaker content II: Subject-matter, common content, remainder and ground. *Journal of Philosophical Logic* 46.6, pp. 675–702. DOI: 10.1007/s10992-016-9419-5.

— (2017c). Truthmaker Semantics. *A Companion to the Philosophy of Language*. Wiley Blackwell. Chap. 22, pp. 556–577. DOI: 10.1002/9781118972090.ch22.

von Fintel, Kai (2001). Counterfactuals in a dynamic context. *Ken Hale: A life in language*. Ed. by Michael Kenstowicz. Vol. 36. MIT Press, pp. 123–152. URL: http://web.mit.edu/fintel/fintel-2001-counterfactuals.pdf.

Hawke, Peter (2018). Theories of aboutness. *Australasian Journal of Philosophy* 96.4, pp. 697–723. DOI: 10.1080/00048402.2017.1388826.

Ippolito, Michela (2013). *Subjunctive Conditionals: A Linguistic Analysis*. MIT Press.

Khoo, Justin (2015). On indicative and subjunctive conditionals. *Philosophers' Imprint* 15.32, pp. 1–40. URL: http://hdl.handle.net/2027/spo.3521354.0015.032.

— (2017). Backtracking counterfactuals revisited. *Mind* 126.503, pp. 841–910. DOI: 10.1093/mind/fzw005.

Kratzer, Angelika (1981). The notional category of modality. *Words, Worlds, and Contexts: New Approaches in Word Semantics. Berlin: W. de Gruyter*. Ed. by Hans-Jürgen Eikmeyer and Hannes Rieser, pp. 39–74. DOI: 10.1515/9783110842524-004.
— (1989). An investigation of the lumps of thought. *Linguistics and philosophy*, pp. 607–653. DOI: 10.1007/BF00627775.
— (2012). *Modals and conditionals*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199234684.001.0001.
Lassiter, Daniel (2018). Complex sentential operators refute unrestricted Simplification of Disjunctive Antecedents. *Semanitcs and Pragmatics* 11. DOI: 10.3765/sp.11.9.
Lewis, David (1973a). Causation. *Journal of Philosophy* 70.17, pp. 556–567. DOI: 10.2307/2025310.
— (1973b). *Counterfactuals*. Wiley-Blackwell.
— (1988). Relevant implication. *Theoria* 54.3, pp. 161–174. DOI: 10.1111/j.1755-2567.1988.tb00716.x.
— (2000). Causation as Influence. *Journal of Philosophy* 97.4, pp. 182–197. DOI: 10.2307/2678389.
Mandelkern, Matthew (2018). Talking about worlds. *Philosophical Perspectives* 32.1, pp. 298–325. DOI: 10.1111/phpe.12112.
McHugh, Dean (forthcoming). Exhaustification in the semantics *cause* and *because*. *Glossa* GLOWing papers: special collection from Generative Linguistics in the Old World.
McKay, Thomas and Peter van Inwagen (1977). Counterfactuals with disjunctive antecedents. *Philosophical Studies* 31.5, pp. 353–356. DOI: 10.1007/BF01873862.
Moss, Sarah (2013). Subjunctive credences and semantic humility. *Philosophy and Phenomenological Research* 87.2, pp. 251–278. DOI: 10.1111/j.1933-1592.2011.00550.x.
Nute, Donald (1975). Counterfactuals and the Similarity of Words. *The Journal of Philosophy* 72.21, pp. 773–778. DOI: 10.2307/2025340.
Pollock, John L (1976). *Subjunctive reasoning*. Springer. DOI: 10.1007/978-94-010-1500-4.
Prior, Arthur Norman (1976). It was to be. *Papers in semantics and ethics*. Ed. by Peter Geach and Anthony Kenny. Duckworth, pp. 97–108.
Reichenbach, Hans (1947). *Elements of symbolic logic*. Dover Publications.
Stalnaker, Robert (1968). A theory of conditionals. *Ifs*. Ed. by William L. Harper, Robert Stalnaker, and Glenn Pearce. Springer, pp. 41–55. DOI: 10.1007/978-94-009-9117-0_2.
Starr, William B. (2014). A Uniform Theory of Conditionals. *Journal of Philosophical Logic* 43.6, pp. 1019–1064. DOI: 10.1007/s10992-013-9300-8.
Thomason, Richmond (1970). Indeterminist time and truth-value gaps. *Theoria* 36.3, pp. 264–281. DOI: 10.1111/j.1755-2567.1970.tb00427.x.
Thomason, Richmond and Anil Gupta (1980). A theory of conditionals in the context of branching time. *Ifs*. Springer, pp. 299–322. DOI: 10.1007/978-94-009-9117-0_15.
Warmbrōd, Ken (1981). Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic* 10.2, pp. 267–289. DOI: 10.1007/BF00248853.
Willer, Malte (2018). Simplifying with Free Choice. *Topoi* 37.3, pp. 379–392. DOI: 10.1007/s11245-016-9437-5.
Yablo, Stephen (2014). *Aboutness*. Princeton University Press. DOI: 10.1515/9781400845989.
Yalcin, Seth (2010). Probability operators. *Philosophy Compass* 5.11, pp. 916–937. DOI: 10.1111/j.1747-9991.2010.00360.x.

An existential semantics for habituals*

Jéssica Mendes

University of Maryland, College Park, MD, USA
jmendes@umd.edu

Abstract

This paper focuses on a new observation concerning the interaction of approximatives (*almost*, *barely*) and habitual sentences: habituals can only be modified by *almost* when negated. Based on the distribution of approximatives with nominal quantifiers, I take this pattern as evidence that habituals have an underlying existential force, and acquire their quasi-universal meaning via scalar strengthening.

1 Introduction

The main empirical contribution of this paper is the observation that, although habituals[1] can be modified by both *almost* and *barely*, the former requires sentential negation. I illustrate this point with data from Brazilian Portuguese (BP), a language that has dedicated habitual morphology.

(2) a. O João quase não checa o email quando tá de férias.
the John almost NEG checks the email when is.3S.HAB on vacation
'John almost doesn't check his email when he's on vacation.'

b. *O João quase checa o email quando tá de férias.
the John almost checks the email when is.3S.HAB on vacation
*'John almost checks his email when he's on vacation.'

c. O João mal checa o email quando tá de férias.
the John barely checks the email when is.3S.HAB on vacation
'John barely checks his email when he's on vacation.'

It's well-known that approximatives place restrictions on their complements. Baron [4], who provides the state-of-the-art treatment of approximatives, formalizes these restrictions in terms of monotonicity: *almost* can only modify operators that are downward-entailing on their

*I'm thankful to my advisors, Paolo Santorio and Valentine Hacquard, for insightful discussions and feedback on this project. I would also like to thank three anonymous reviewers for the Amsterdam Colloquium, attendees of the Meaning Meeting at UMD, Christopher Baron and Omar Agha for their helpful comments. All errors are my own.

[1] There is a growing consensus in the literature that constructions that have been historically labeled 'habituals' don't form a unified class. In recent work, Cable [5] distinguishes between habituals that express unrealized dispositions (like *Mary handles the mail from Antarctica*), and 'actualized habituals', which describe habits that have been realized in the actual world. It seems that *barely* and *almost-not* can only modify the latter:

(1) a. ?Mary almost doesn't handle the mail from Antarctica.
b. ?Mary barely handles the mail from Antarctica.

This speaks in favor of Cable's [5] hypothesis that purely dispositional habituals and actualized habituals involve different LFs even in languages that don't show this distinction overtly.

left argument, while *barely* requires left upward-entailing operators.[2] These constraints neatly capture the complementary distribution of *almost* and *barely* with nominal quantifiers, shown in (3).

(3) a. Almost every student came to the meeting. ✓*almost* ∀
b. Almost no student came to the meeting. ✓*almost* ¬∃
c. *Almost some student came to the meeting. **almost* ∃
d. Barely any student came to the meeting. ✓*barely* ∃
e. *Barely every student came to the meeting. **barely* ∀

In light of these facts, the paradigm in (2) provides evidence that habitual sentences involve existential quantification. This conclusion is rather puzzling: intuitively, habituals seem to have (quasi-)universal force. Before discussing how these seemingly conflicting facts can be reconciled, let me say a word about the broader significance of this observation.

Habitual sentences give rise to Homogeneity effects, as first noticed by Ferreira [9]. That is, a bare habitual, like (4a), has a universal reading, but its negated form (4b) has the force of a negated existential. Habituals are also known to tolerate exceptions—(4a) could be truthfully uttered even if John doesn't *always* smoke when he's anxious; exception tolerance, or non-maximality, is a hallmark of constructions that display Homogeneity. Finally, a universal quantifier can remove the Homogeneity of (4a); unlike (4a), (4c) doesn't tolerate exceptions to John's behavior.

(4) a. John smokes when he's anxious.
i. ≈John *always* smokes when he's anxious.
ii. ≉John *sometimes* smokes when he's anxious.
b. John doesn't smoke when he's anxious.
i. ≉John doesn't *always* smoke when he's anxious.
ii. ≈John *never* smokes when he's anxious.
c. John *always* smokes when he's anxious.

These facts find analogs in the behavior of definite plurals, the most well-understood example of Homogeneity effects in natural language:

(5) a. The semantics students came to the meeting.
i. ≈*Every* semantics student came to the meeting.
ii. ≉*Some* semantics students came to the meeting.
b. The semantics students didn't come to the meeting.
i. ≉Not *every* semantics student came to the meeting.
ii. ≈*No* semantics student came to the meeting.
c. *Every* semantics student came to the meeting.

It seems likely that habituals and definite plurals are simply different exponents of Homogeneity, and thus deserve a unifying account. The observation in (2) adds a desideratum to this account: it should include an operator that is upward monotonic on its restrictor. To my knowledge, the only existing account that meets this criteria is Bar-Lev [3], according to which Homogeneity effects arise when an existential distributive operator (∃-DIST) undergoes

[2] I'm omitting some complications posed by *barely* that are not immediately relevant to the discussion in this paper. I refer the reader to Baron's dissertation for a detailed explanation of how his account can be extended to other uses of approximatives, including cases in which they modify adjectives and numerals.

strengthening. The paradigm in (2) provides support for the presence of this operator in the syntax. Therefore, the discussion in this paper may not only contribute to our understanding of habituals, but also to the debate about the proper derivation of Homogeneity effects in natural language.

2 Proposal

2.1 An existential semantics for habituals

As I mentioned above, the facts in (2) can be straightforwardly captured by extending Bar-Lev [3]'s treatment of definite plurals to habituals. A crucial feature of this account is that sentences with definite plurals have a weak, existential semantics, contributed by the distributive operator $\exists$-DIST$_D$, which is obligatorily adjoined to the sister of plural DPs. A sentence like (6a) has the LF in (6b):

(6) a. The girls danced.
b. [The girls][$\exists$-DIST$_D$ danced]

Definite plurals are non-quantificational and simply denote the maximal element in a plurality. That is, if there are two girls under discussion, say, Anna and Mary, $[\![\text{the girls}]\!] = \text{Anna} \sqcup \text{Mary} = \text{Anna} \vee \text{Mary}$. The role of definite plurals in this analysis is to restrict the domain of the variable D to the atomic parts of the plurality it denotes. $\exists$-DIST$_D$ has the lexical entry in (7a).

(7) a. $[\![\exists\text{-DIST}_D]\!] = \lambda P.\lambda x.\exists y \in D \cap Part_{AT}(x)[P(y) = 1]$
b. $Part_{AT}(x) = \{y : y \sqsubseteq_{AT} x\}$

The final meaning of (6a) is obtained via scalar strengthening, which is carried out by a covert exhaustification operator, whose lexical entry is given below.

(8) $[\![\text{EXH}]\!](C)(p) \Leftrightarrow \forall q \in IE(p, C)[\neg q] \wedge \forall r \in II(p, C)[r]$

When applied to an alternative set C and a proposition p, EXH negates all innocently excludable (IE) alternatives, and asserts all innocently includable (II) alternatives to p in C. IE alternatives are formed by intersecting all maximal subsets of C that can be negated consistently with the prejacent. The II alternatives are formed by the intersection of all maximal subsets of C that can be asserted consistently with the prejacent and with the negation of the IE alternatives. This complex algorithm is designed to ensure that exhaustification does not lead to a contradiction, and that all alternatives in C receive a truth value whenever possible.

With the basic ingredients of Bar-Lev [3]'s account on the table, we're ready to switch our focus back to habituals. As we saw, in sentences like (6a), definite plurals act as restrictors to the domain of $\exists$-DIST$_D$. What could be playing the same role in sentences with habituals? Non-quantificational temporal adverbials, like *on Fridays*, and clauses headed by *when*, *before* and *after* seem to be good candidates:

(9) a. John smokes when he's anxious.
b. John meditates before a meeting.
c. John smokes after drinking coffee.
d. John exercises on Fridays.

Of course not every habitual sentence includes an overt restrictor like the ones above. Very often, this restrictor is recovered from the context. For example, B's reply in (10) would be

naturally understood as being restricted to the times when B watches the World Cup. In some cases, it's not clear that there is a restrictor at all. I delay a discussion of these cases to §4.

(10) A: Are you gonna watch the World Cup?
B: No, I get too emotionally invested.

For concreteness, I assume the simplified LF in (11) for (9d).

(11) [On Fridays][$\exists$-DIST$_D$ John exercises]

To pursue the analogy with Bar-Lev's [3] account of definite plurals, I assume non-quantificational temporal adverbials, like 'on Fridays' denote a plurality of times. I furthermore adopt a toy model containing only two Fridays, so $[\![\text{on Fridays}]\!] = i_1 \vee i_2$. The alternatives to (9d) are obtained by replacing the domain variable D with its subsets.

(12) a. $Alt([\![9d]\!]) = \{[\text{on Fridays}][\exists\text{-DIST}_{D'} \text{ John exercises}] : D' \subseteq D\}$
b. $= \{\text{J exercises at } i_1 \vee \text{J exercises at } i_2, \text{J exercises at } i_1, \text{J exercises at } i_2\}$

Notice (12b) does not contain the conjunctive alternative {J exercises at $i_1 \wedge$ J exercises at i_2}. The assumption that (12b) is not closed under conjunction is crucial to deriving strengthening, as it renders all alternatives in (12b) *II*. As a result, all alternatives are asserted, and $\exists$-DIST$_D$ acquires a universal force.

(13) $[\![\text{EXH}]\!](Alt([\![6a]\!]))([\![6a]\!]) = \text{J exercises at } i_1 \vee \text{J exercises at } i_2$
$\wedge$ J exercises at $i_1 \wedge$ J exercises at i_2
$= \text{J exercises at } i_1 \wedge \text{J exercises at } i_2$

Applying EXH above negation has no truth-conditional effect, as the negated prejacent $\neg$(J exercises at $i_1 \vee$ J exercises at i_2) is already the strongest alternative. That correctly captures the contrast between (5a) and (5b), which is the hallmark of Homogeneity.[3]

The discussion above might raise a concern related to the scope of approximatives and EXH. Recall from §1 that *almost* modifies universal quantifiers and negated existentials, while *barely* modifies only existentials. Now, suppose EXH intervenes between *almost* and $\exists$-DIST$_D$, as shown in (14a). $\exists$-DIST$_D$ would be strengthened into a universal meaning via exhaustification before combining with the approximatives. It is tempting to conclude that, in this configuration, *almost* and *barely* would 'see' $\exists$-DIST$_D$ as a universal, which, given the paradigm in (4), never seems to be the case.[4]

(14) a. *[*almost* [EXH [$\exists$-DIST ...]]]
b. [EXH [*almost* [$\exists$-DIST ...]]]

As I said earlier, Baron [4] states the selectional requirements of approximatives in terms of monotonicity: *almost* can only modify left downward-entailing operators. Even if $\exists$-DIST was strengthened before combining with *almost*, it would still have the formal properties of

[3] In Bar-Lev [3]'s system, non-maximal readings are derived by capitalizing on the context-sensitive nature of implicature calculation. For reasons of space, I cannot discuss these readings here, but Bar-Lev's derivation of non-maximality with definite plurals should carry over straightforwardly to habituals.

[4] This discussion focuses only on *almost* because *barely* encodes negation in its lexical entry (see Baron, [4]). EXH cannot apply under negation (Fox & Spector, [10]), so an LF in which EXH scoped under *barely* would be ruled out on independent grounds.

an existential—after all, strengthening consists in conjoining an existential prejacent with its sub-domain alternatives. I conclude that, if the LF in (14a) is available at all, it leads to trivial truth conditions, and therefore, ungrammaticality.[5]

3 Comparison with previous accounts

I take it that the primary goal of any theory of habituals should be a unification with definite plurals. Based on the interaction of habituals and approximatives, I argued that a theory of habituals must also give them an underlying existential force. In this section, I compare my account to Ferreira [9] and Agha [1], the two existing analyses of habituals that are explicit about their connection with definite plurals.

Ferreira, [9], who first noticed this link, took some important steps towards a unification of the two phenomena, but ultimately fell short of his goal. Ferreira argues that the restrictors of habituals are, themselves, universal distributive operators. Then, to capture the exception-tolerance of habituals, Ferreira [9] needs to resort to a modal-temporal operator (IMP) which the author takes to be present in progressive and habitual sentences.[6] This move is undesirable because, as we saw, definite plurals are also tolerant to exceptions—and they're unlikely to involve any modality.

Agha [1] offers an elegant treatment of habituals based on Križ's [13]'s account of Homogeneity. His proposal shares a number of similarities with the present account. In brief, the author argues that habitual readings are obtained when a sentence radical, which he takes to denote a predicate of times, is applied directly to a tense head or frame adverbial, which denote a plurality of time intervals. One of Agha's goals with his non-quantificational account is to avoid the undesirable consequences of giving habituals a universal meaning. In this respect, our views are very much aligned. However, since Agha's semantics for habituals doesn't involve any quantification, it's not clear how it could capture the pattern in (2).

4 Conclusion & open issues

This paper brought new evidence that suggests that habituals have an underlying existential force. I then showed that this otherwise puzzling fact can be neatly accounted for by adopting Bar-Lev's [3] view of Homogeneity as scalar strengthening.

To push the analogy between habituals and definite plurals, I assumed the LF of habituals always includes a restrictor, which can be overt or recovered from the context. This assumption is shared by other accounts, like Ferreira's and Cable's [5]. However, some habitual sentences don't seem to involve a restrictor at all; (16) could be uttered in an out-of-the-blue context to convey that John is a smoker.

(16) John smokes.

[5] Free Choice Items, like *any* in some of its uses, are often treated as existentials that undergo strengthening (see Chierchia, [6], and Dayal, [7], among many others). Free Choice *any* can, however, be modified by *almost*:

(15) You can pick almost any card.

If the conclusion of this section is on the right track, it suggests that, contrary to common wisdom, FCI *any* is a universal. An in-depth discussion of this contentious point is left for future work.

[6] Ferreira's modal treatment of habituals is inspired by Portner's [15] semantics for the progressive.

We could assume examples like (16) include a trivial restrictor ('John smokes [when he smokes]'), but there's evidence that these cases deserve a separate treatment.[7] Habituals like (16) behave differently from the other habituals discussed in this paper in at least one more way: (16) gives rise to the 'same-participant effect' (as noted by Krifka et al. [12], Ferreira [9], [8], a.o.). Unlike (18a) and (18b), (18c) sounds odd because it implies that John smokes the same cigarette multiple times. Integrating examples like (16) into my account is left for future work.

(18) a. When he's anxious, John smokes a cigarette.
b. John smokes a cigarette after dinner.
c. #John smokes a cigarette.

Another issue raised by my proposal concerns some disanalogies between definite plurals and habituals. If these two constructions have such similar semantics, all else equal, they should also behave the same way with respect to modification by *almost* and *barely*. Before addressing this concern, I should point out that my proposal does *not* predict that *almost* or *barely* should be able to attach directly to the definite plural, as in (19).

(19) *Barely the girls danced.

It's fairly uncontroversial to assume that, despite their cross-categorial nature, *almost* and *barely* impose restrictions on the semantic type of their prejacents. More specifically, it seems that *almost* and *barely* take prejacents of a high type, like quantifiers. Definite plurals are normally not assumed to have such high types (see Link [14], Krifka [11], a.o.), so (19) could be ruled out on the basis of a type mismatch.

Examples (20a), (20b) and (20c) deserve more attention. Assuming *almost* and *barely* scope over $\exists$-DIST$_D$ in these sentences, my account predicts (20a) and (20b) should be acceptable and truth-conditionally equivalent, conveying that very few girls danced. It also predicts, all else being equal, that (20c) should be unacceptable. These predictions are not borne out.

(20) a. The girls barely danced.
b. The girls almost didn't dance.
c. The girls almost danced.

Let me start by addressing the acceptability of (20c). It seems that the only available reading of (20c) is counterfactual; (20c) could be paraphrased as 'all girls came close to dancing, but didn't.' Amaral [2] and Rapp & von Stechow [16], a.o., point out that counterfactual readings of *almost* are only possible in sentences with certain mood and aspectual morphemes. This observation suggests that in (20c), *almost* is not modifying $\exists$-DIST$_D$, but rather some intensional operator whose presence is obscured in English. Examples (20a) and (20b), on the other hand, seem to give rise to a reading that could be paraphrased as 'the girls danced very little'. It's not clear which operator *almost-not* and *barely* are modifying in this case. I must leave a more detailed comparison between habituals and definite plurals to future work.

[7] Notice, however, that (16) patterns with other habituals with respect to modification by *almost* and *barely*, which could be taken as evidence that (16) is not a 'dispositional habitual' (see fn.1):

(17) a. *John almost smokes.
b. John almost doesn't smoke.
c. John barely smokes.

References

[1] Omar Agha. Temporal homogeneity and exceptions to habituals. Paper presented at Semantics and Linguistic Theory 31, online, 2021.

[2] Patrícia Amaral and Fabio Del Prete. Approximating the limit: the interaction between *quasi* 'almost' and some temporal connectives in Italian. *Linguistics and Philosophy*, 33(2):51–115, 2010.

[3] Moshe E Bar-Lev. An implicature account of homogeneity and non-maximality. *Linguistics and Philosophy*, 44(5):1045–1097, 2021.

[4] Christopher Baron. *The logic of subtractives or Barely anyone tried almost as hard as me*. PhD thesis, Massachusetts Institute of Technology, 2022.

[5] Seth Cable. Two paths to habituality: The semantics of habitual mode in Tlingit. *Semantics and Pragmatics*, 15:11–EA, 2022.

[6] Gennaro Chierchia. *Logic in grammar: Polarity, Free Choice, and intervention*. OUP Oxford, 2013.

[7] Veneeta Dayal. A viability constraint on alternatives for Free Choice. In *Alternatives in semantics*, pages 88–122. Springer, 2013.

[8] Marcelo Ferreira. The semantic ingredients of imperfectivity in progressives, habituals, and counterfactuals. *Natural Language Semantics*, 24(4):353–397, 2016.

[9] Marcelo Barra Ferreira. *Event quantification and plurality*. PhD thesis, Massachusetts Institute of Technology, 2005.

[10] Danny Fox and Benjamin Spector. Economy and embedded exhaustification. *Natural Language Semantics*, 26(1):1–50, 2018.

[11] Manfred Krifka. Definite NPs aren't quantifiers. *Linguistic Inquiry*, 23(1):156–163, 1992.

[12] Manfred Krifka, Francis Jeffry Pelletier, Gregory Carlson, Alice Ter Meulen, Gennaro Chierchia, and Godehard Link. Genericity: an introduction. In *The Generic Book*. 1995.

[13] Manuel Križ. *Aspects of homogeneity in the semantics of natural language*. PhD thesis, University of Vienna, 2015.

[14] Godehard Link et al. The logical analysis of plurals and mass terms: A lattice-theoretical approach. *Formal semantics: The essential readings*, 127:147, 1983.

[15] Paul Portner. The progressive in modal semantics. *Language*, pages 760–787, 1998.

[16] Irene Rapp and Arnim Von Stechow. *Fast* 'almost' and the visibility parameter for functional adverbs. *Journal of Semantics*, 16(2):149–204, 1999.

Farsi *Hame -i* DPs and Intensional Transitive Verbs*

Luis Alonso-Ovalle[1] and Esmail Moghiseh[2]

[1] McGill University, Montréal, Québec, Canada
luis.alonso-ovalle@mcgill.ca
[2] McGill University, Montréal, Québec, Canada
esmail.moghiseh@mail.mcgill.ca

Abstract

Adverbial modification suggests that some, but not all, Intensional Transitive Verbs (ITVs) take clausal complements [11]. Evidence from Farsi, coming from the interpretation of *hame -i* DPs, confirms the split. *Hame -i* DPs are existential quantifiers that acquire universal force via exhaustification [6]. We provide new data showing that *hame -i* DPs can have free choice interpretations, but only under certain conditions. The pattern is new. Free choice interpretations of *hame -i* DPs are found with ITVs, but only those that do not seem to take clausal complements. ITVs that seem to take clausal complements pattern with modal auxiliaries, which don't allow for free choice readings. If free choice readings are derived when a modal expression intervenes between an exhaustivity operator, EXH, and the DP, this means that not all modal expressions can intervene between EXH and *hame -i* DPs. To account for the restricted availability of free choice interpretations, we propose that EXH and *hame -i* DPs need to be within the same minimal clause.

1 Introduction

Hame -i DPs are Farsi DPs that range over types of individuals. As we will see in Section 2, these DPs can have existential or universal force. In [6] we analyze *hame -i* DPs as existential quantifiers and derive their universal force through exhaustification. The main goal of this paper is to document, in Sections 3 and 4, that *hame -i* DPs can have free choice interpretations, but only with modal expressions that do not take clausal complements. We tentatively propose in Section 5 that this behavior follows from a locality constraint requiring the exhaustification operator and *hame -i* DPs to be within the same minimal clause.

2 *Hame -i* DPs: from existentials to universals

In Farsi, existential quantification can be expressed with bare NPs, as shown in (1), or with NPs marked with enclitic *-i*, as seen in (2).

(1) Forood too chale oftad.
Forood in hole fell.3SG
'Forood fell in a hole/holes.'

(2) Forood too chale-i oftad.
Forood in hole-IND fell.3SG
'Forood fell in a hole.'

A number of determiners combine with bare NPs or with NPs marked with *-i* to form DPs that still convey existential quantification. The existential DPs in (3) and (4), for instance, feature numeral *ye* ('one').

*Thanks to the anonymous reviewers for the Amsterdam Colloquium. Funding: Social Sciences and Humanities Research Council of Canada, Insight Grant (*Modality across Categories*, 435-2018-0524, PI: Alonso-Ovalle). Our names are listed in alphabetical order.

(3) Forood too ye chale oftad.
Forood in one hole fell.3SG
'Forood fell in a hole.'

(4) Forood too ye chale-i oftad.
Forood in one hole-IND fell.3SG
'Forood fell in a hole.'

Other determiners can combine with NPs marked with *-i* [4, 5]. In positive episodic sentences with non-volitional verbs, adding *hame* to an NP marked with *-i* conveys universal quantification over *types* of entities: (5), for instance, conveys that for *each* (contextually relevant) type of (contextually relevant) hole P, Forood fell in a hole of type P.[1]

(5) Forood too hame chale-i oftad.
Forood in hame hole-IND fell.3SG
'Forood fell in all types of holes.'

The universal force of *hame -i* DPs can turn to existential in a downward entailing (DE) environment: (6), for instance, conveys that few students fell in some type of hole or other [6].

(6) Tedad-e kam-i danshjoo too hame chale-i oftad-and.
number-EZ little-IND student in hame hole-IND fell-3PL
'Few students fell in any type of hole.'

In [6] we propose to capture the shift in quantificational force between positive episodic and DE environments by assuming that *hame -i* DPs have existential force that can be strengthened to universal via exhaustification, in line with the analysis of the Hebrew determiner *kol* presented in [8]. We take *hame -i* DPs to be existentials that quantify over cells of a partition, as in (7-a) (where $\mathbb{C}$ retrieves a contextually relevant set of properties that partitions (a contextually relevant) set of holes) and assume that *hame -i* DPs introduce into the semantic derivation domain but not scalar alternatives.[2] The domain alternatives that *hame -i* DPs introduce are determined by considering subsets of the original partition of the relevant domain of entities in the extension of the NP, as in (7-b).

(7) a. $[\![\text{hame}_{\mathbb{C}}\ \text{hole-i}_C]\!]^{\text{o}} = \lambda Q.\exists P \in \mathbb{C}([\![\text{hole}_C]\!])\exists x[[\![\text{hole}_C]\!](x) \wedge P(x) \wedge Q(x)]$
b. $[\![\text{hame}_{\mathbb{C}}\ \text{hole-i}_C]\!]^{\text{ALT}} = \{\lambda Q_{et}.\exists P \in D \exists x[[\![\text{hole}_C]\!](x) \wedge P(x) \wedge Q(x)] : D \subset \mathbb{C}([\![\text{hole}_C]\!])\}$

As is standard, we assume that these alternatives grow propositional and are accessed by an exhaustification operator at propositional sites. To illustrate, consider the LF for (5) in (8). Assuming that the contextually restricted extension of the NP consists of four holes, of two different colours, red and green, and that $\mathbb{C}$ retrieves the property of being red and the property of being green, as in (9), the sister of EXH in (8) (its 'prejacent') denotes the disjunction in (10) at the ordinary semantic level and the set in (11) at the alternative level.[3]

[1] We restrict our attention to cases with non-volitional verbs in positive episodic environments because with volitional verbs *hame -i* DPs *can* (but don't have to) convey a 'random choice' interpretation [2]. The sentence in (i), for instance, can convey that Forood bought *a* book of *some* type and that he chose the type randomly. These cases may involve covert modality [9], and, so, we leave them aside for now. Note, however, that if these cases involved covert modality and the covert modal did not take a clausal complement, they would also fall under the proposal presented in the last section of the paper.

(i) Forood hame book-i xarid.
Forood hame book-IND bought.3SG

[2] The types that *hame -i* DPs range over are not necessarily stable, hence our choice is to resort to quantification over properties rather than subkinds. We nevertheless leave open for now the possibility that *hame -i* DPs range over *ad hoc* subkinds.

[3] From now on, for the purposes of illustration, we will represent each domain alternative as 'G' and 'R'.

(8) LF: EXH [$_{IP}$ hame$_{\mathbb{C}}$ hole-i$_C$ λ_1 Forood fell in t_1]

(9) $[\![\text{hole}_C]\!] = \{R_1, R_2, G_1, G_2\}$, $\mathbb{C}([\![\text{hole}_C]\!]) = \{\text{RED}, \text{GREEN}\}$

(10) $[\![[_{IP}\ \text{hame}_{\mathbb{C}}\ \text{hole-i}_C\ \lambda_1\ \text{Forood fell in}\ t_1]]\!]^O =$
$\exists x[\text{HOLE}(x) \wedge \text{RED}(x) \wedge \text{FALL}(F,x)] \ \vee\ \exists x[\text{HOLE}(x) \wedge \text{GREEN}(x) \wedge \text{FALL}(F,x)]$

(11) $[\![[_{IP}\ \text{hame}_{\mathbb{C}}\ \text{hole-i}_C\ \lambda_1\ \text{Forood fell in}\ t_1]]\!]^{ALT} = \left\{ \begin{array}{c} \exists x[\text{HOLE}(x) \wedge \text{RED}(x) \wedge \text{FALL}(F,x)], \\ \exists x[\text{HOLE}(x) \wedge \text{GREEN}(x) \wedge \text{FALL}(F,x)] \end{array} \right\}$

EXH strengthens its prejacent by negating all 'innocently excludable' alternatives and asserting it together with all 'innocently includable' alternatives. An alternative is innocently excludable if it's in every set containing the prejacent and as many negated alternatives as consistency allows for. An alternative is innocently includable if it can be asserted together with the prejacent and all negated innocently excludable alternatives (if there are any) [7]. The set containing all maximal subsets of alternatives whose negation is consistent with (10) together with (10) is in (12). There is no alternative that is in all members of (12), and, therefore, there is no innocently excludable alternative. All alternatives are innocently includable. Conjoining the disjunction in (10) with all alternatives turns the disjunction into a conjunction, thus deriving the attested universal force, as seen in (13).

(12) $\{\{(10), G\}, \{(10), R\}\}$

(13) $[\![(8)]\!]^O = (G \vee R) \wedge G \wedge R \Leftrightarrow G \wedge R$

In DE contexts, like (14), global exhaustification is vacuous. We see in (15) that the prejacent entails now all the alternatives, so none is innocently excludable. All alternatives are innocently includable, but conjoining them with the prejacent has no effect, since they are entailed by it.

(14) Age Forood too hame chale-i bioft-e, shart ro mibar-am.
if Forood in hame hole-IND fall-3SG, bet ACC win-1SG
'If Forood falls in some type of hole or other, I win the bet.'

(15) a. EXH [$_{IP}$ if hame$_{\mathbb{C}}$ hole-i$_C$ λ_1 Forood fall in t_1, I win the bet]
b. $[\![[_{IP}\ \ldots\ ]]\!]^O = (G \vee R) \rightarrow \text{BET}$, $[\![[_{IP}\ \ldots\ ]]\!]^{ALT} = \{G \rightarrow \text{BET}, R \rightarrow \text{BET}\}$
c. $[\![\text{EXH}\ [_{IP}\ \ldots\ ]]\!]^O = [(G \vee R) \rightarrow \text{BET}] \wedge G \rightarrow \text{BET} \wedge R \rightarrow \text{BET} \Leftrightarrow (G \vee R) \rightarrow \text{BET}$

Assuming that *hame -i* DPs are existential quantifiers that introduce only domain alternatives naturally captures the oscillation in quantificational force that we have attested.

3 Interaction with modals

The existential force of *hame -i* DPs is not limited to DE contexts. We also find it in some (but not all) modal contexts. Intensional Transitive Verbs (ITVs) are a case at hand. Consider, for instance, (16), with the Farsi counterpart of *look for*, which is expressed with the 'light' verb *gashtan* ('to search') and the noun *donbal* ('pursuit') ('search in pursuit of.')

(16) Forood donbal-e hame ketab-i gasht.
Forood pursuit-EZ hame book-IND searched.3SG
'Forood looked for some type of book or other.'

The truth of (16) does not require that Forood looked for *every* type of book: (16) is true in the scenario in (17). The sentence conveys a free choice effect: it is true just in case Forood wanted to have *some* type of book or other and *any type* was a good option for him.

(17) There are two types of books: red and green. Forood didn't look for red books and green books. He looked for a book of one of these colors: it could be red or green.

Within the framework that we are assuming, free choice readings are derived from LFs where a modal expression intervenes between EXH and the existential *hame -i* DP, as in (18-a). Assuming that $\mathbb{C}([\![\text{book}_C]\!]) = \{\text{G}, \text{R}\}$, the prejacent of EXH in (18-a) conveys the proposition in (18-b). All alternatives, in (18-b), are now innocently excludable and none are innocently includable. EXH excludes all alternatives, as in (18-c), yielding a free interpretation: that Forood wanted to have a book of some type and any type was a good option for him.

(18) a. $\text{EXH}\ \Box\ [_{\text{IP}}\ \text{hame}_{\mathbb{C}}\ \text{hole-i}_C\ \lambda_1\ \text{Forood bought}\ t_1]$
b. $[\![\Box\ [_{\text{IP}}\ \ldots\ ]]\!]^{\text{O}} = \Box(\text{G} \vee \text{R}),\ [\![\Box\ [_{\text{IP}}\ \ldots\ ]]\!]^{\text{ALT}} = \{\Box\text{G},\ \Box\text{R}\}$
c. $[\![\text{(18-a)}]\!] = \Box(\text{G} \vee \text{R}) \wedge \neg\Box\text{G} \wedge \neg\Box\text{R} \Leftrightarrow \Box(\text{G} \vee \text{R}) \wedge \Diamond\text{G} \wedge \Diamond\text{R}$

Hame -i DPs do not have free choice readings with all ITVs, though. The sentence in (19), with the Farsi counterpart of *want*, for instance, cannot describe the scenario in (17). The sentence conveys that for each (contextually relevant) type of (contextually relevant) book P, Forood wanted a book of type P.

(19) Forood hame ketab-i khast.
Forood hame book-IND wanted.3SG
'Forood wanted all types of books.'

Modal auxiliaries pattern with *want*, rather than with *look for*. To illustrate, consider the sentence in (20) below, with the necessity modal auxiliary *bayad* ('must'). The *hame -i* DP in (20) has universal force. The sentence conveys that for each type of book P, Forood is required to buy a book of type P. The sentence is therefore false in the context in (21), where Forood is not required to buy books of all types. In (21), the potential free choice interpretation is true, but the sentence is false.[4,5]

(20) Forood bayad hame ketab-i bekhar-e.
Forood must hame book-IND buy-3SG
'Forood must buy all types of books.'

(21) There are two types of books (red, green). Forood is required to buy a book of some type or other—any type. He is not required to buy red books and green books.

We conclude that *hame -i* DPs do not convey a free choice interpretation with all expressions. Free choice interpretations are known to be limited to some modal flavors [1], but the data

[4]One may think that the contrast between the Farsi counterparts of *must* and *look for* is due to the complex structure of the latter. We can see that this is not the case with the help of non-modal complex verbs that have a construction similar to that of *look for*: (i), with *savar shodan* ('to get on'), composed of the noun *savar* ('passenger') and the light verb *shodan* ('to become'), does not convey free choice. The sentence conveys that for each type of car P, Forood is required to get on a car of type P.

(i) Forood bayad savar-e hame mashin-i beshe-e.
Forood must passenger-EZ hame car-IND become-3SG

[5]The sentence in (i), with a possibility modal auxiliary, conveys that for each type of book P, Forood is allowed to buy a book of type P. Notice, however, that this can be due to a wide scope universal interpretation of the *hame -i* DP, hence our focus on the necessity modal auxiliary case.

(i) Forood mitun-e hame ketab-i bekhar-e.
Forood can-3SG hame book-IND buy-3SG

pattern that we illustrate here seems new: in the next section, we point out that there is a correlation between the types of modal expressions that allow for a free choice interpretation with *hame -i* DPs and the possible syntactic size of their complements.

4 A correlation between types of ITVs and FC readings

In [10], Barbara Partee noted a split between types of ITVs that has to do with the interpretation of adverbial modifiers. Consider the case of *want*: (22-a) can be paraphrased as in (22-b). In (23-a) we add the temporal modifier *by Saturday*. The interpretation of (23-a) can be paraphrased with (23-b), where the adverbial phrase is part of the clausal complement.

(22) a. Martha wants an apartment.
b. Mary wants [to have an apartment].

(23) a. Martha wants an apartment by Saturday.
b. Mary wants [to have an apartment by Saturday].

Now consider (24-a), with *look for*, which can be paraphrased with (24-b). We could then entertain the possibility of reducing *look for* to *trying to find*. If so, we would expect the temporal modifier in (25-a) to be able to be interpreted within the complement clause of *try*, and for (25-a) to be possibly paraphrased with (25-b). This is not the case: the temporal adverbial *by Saturday* can only convey information about the search, not the goal of the search.

(24) a. Martha is looking for an apartment.
b. Martha is trying [to find an apartment]

(25) a. Martha is looking for an apartment by Saturday.
b. Martha is trying [to find an apartment by Saturday.]

Based on contrasts of this nature, Schwarz [11] suggests it is plausible to assume that some but not all ITVs take a hidden clausal complement.

The contrast between (23-a) and (25-a) holds between the Farsi counterparts of *want* and *look for*, too. To illustrate: (26-a), with *khastan* ('want'), can be paraphrased with (26-b), where *want* takes a full clausal complement which the adverbial *dirooz* ('yesterday') is part of. This contrasts with the sentence in (27-a), with the Farsi counterpart of *look for*. The sentence in (27-a) sounds contradictory because the two adverbials *ye hafte pish* ('a week ago') and *dirooz* ('yesterday') convey contradictory information. This contrasts with the sentence in (27-b), where *dirooz* ('yesterday') is interpreted within the clausal complement of *try* and conveys information about Forood's goal.

(26) a. Ye hafte pish, Forood ye mashin dirooz khast.
one week ago, Forood one car yesterday wanted.3SG
'A week ago, Forood wanted a car yesterday.'
b. Ye hafte pish, Forood ye mashin dirooz khast dashte bash-e.
one week ago, Forood one car yesterday wanted.3SG have be-3SG
'A week ago, Forood wanted to have a car yesterday.'

(27) a. #Ye hafte pish, Forood donbal-e ye mashin dirooz gasht.
one week ago, Forood pursuit-EZ one car yesterday searched.3SG
'A week ago, Forood looked for a car yesterday.'

b. Ye hafte pish, Forood talash kard ye mashin dirooz payed kon-e.
one week ago, Forood try did.3SG one car yesterday find do-3SG
'A week ago, Forood tried to find a car yesterday.'

Necessity modal auxiliaries take clausal complements: as expected, (28) is not contradictory.

(28) Ye hafte pish, Forood bayad ye mashin dirooz mikharid.
one week ago, Forood must one car yesterday bought.3SG
'A week ago, Forood was required to buy a car yesterday.'

We are then left with a correlation: the ITVs that allow for *hame -i* DPs to have a free choice reading are those that do not seem to take clausal complements. The free choice interpretation of *hame -i* DPs supports the split between ITVs that the adverbial modification facts motivate.

5 A tentative conclusion

Recall that within the framework that we are assuming, free choice readings are derived from LFs where a modal expression intervenes between EXH and the *hame -i* DP. In view of this, we tentatively conclude that the correlation attested in the previous two sections illustrates that the relation between EXH and *hame -i* DPs is subject to a locality constraint. We suggest, in (29), that EXH and *hame -i* DPs need to be within the same minimal clause at LF.

(29) *LF: EXH $\Box$ [$_{CP}$ ... hame$_{\mathbb{C}}$ hole-i$_C$ ...]

Modals that do not take clausal complements, like the Farsi counterpart of *look for*, can intervene between EXH and the *hame -i* DP without violating the locality constraint in (29), and free choice readings are derived. For modals taking clausal complements, like the Farsi counterpart of *want* or *must*, *hame -i* DPs cannot remain below the modal and no free choice interpretation is derived. In those cases, *hame -i* DPs need to be extracted over the modal to yield a licit LF, as in (30). This correlates with the attested universal force [6].

(30) LF: EXH [hame$_{\mathbb{C}}$ hole-i$_C$] λ_1 $\Box$ [$_{CP}$ ... t_1]

While consistent with the free choice interpretation pattern, (29) is at odds with our description of the behavior of *hame -i* DPs in DE contexts, since in Section 2 we took the embedded existential reading of *hame -i* DPs to arise from an LF of the form in (31), which violates (29).

(31) LF: EXH [$_{CP}$ [$_{CP}$ if ... [hame$_{\mathbb{C}}$ hole-i$_C$] ...] ...]

In [6] we show that *hame -i* DPs can be locally exhaustified in DE environments, as in (32). LFs like (32) obey the locality constraint in (29). In cases like (32), *hame -i* DPs contribute universal force within the minimal clause that they are contained in. But that's not the only possibility. In [3] we defend that the alternatives introduced by NPs marked with *-i* can be deactivated under the threat of deriving a contradiction. If we assume that the alternatives can be optionally pruned or deactivated in (32), we can derive the attested narrow scope existential force in DE environments via an LF that obeys the locality constraint.

(32) LF: [$_{CP}$ [$_{CP}$ if EXH [... [hame$_{\mathbb{C}}$ hole-i$_C$] ...] ...] ...]

To avoid overgeneration of existential readings, we need to assume that total pruning or deactivation is only allowed when it leads to global strengthening.

References

[1] Luis Alonso-Ovalle and Paula Menéndez-Benito. Epistemic indefinites: An overview. In Luis Alonso-Ovalle and Paula Menéndez-Benito, editors, *Epistemic Indefinites*, pages 1–30. Oxford University Press, 2015.

[2] Luis Alonso-Ovalle and Paula Menéndez-Benito. Projecting possibilities in the nominal domain: Spanish *uno cualquiera*. *Journal of Semantics*, 35(1):1–41, 2018.

[3] Luis Alonso-Ovalle and Esmail Moghiseh. Contradiction-free strengthening and alternative discharge: The case of Farsi *-i* indefinites. In María Teresa Espinal et al., editor, *Sinn und Bedeutung (SuB) 23*, pages 1–18. Universitat Pompeu Fabra, Universitat Autònoma de Barcelona and Ikerbasque and UPV/EHU, Universidad de Alcalá and Universitat Pompeu Fabra, 2019.

[4] Luis Alonso-Ovalle and Esmail Moghiseh. Neutralizing free choice items via maximal domain restriction: Farsi *i-* indefinites. In *Semantics and Linguistic Theory (SALT) 29*, volume 29, pages 686–705, 2019.

[5] Luis Alonso-Ovalle and Esmail Moghiseh. Numeral *any*: the view from Farsi. In *Semantics and Linguistic Theory (SALT) 30*, volume 30, pages 485–503, 2021.

[6] Luis Alonso-Ovalle and Esmail Moghiseh. Universal force from exhaustification: Farsi *hame -i* DPs. In *Semantics and Linguistic Theory (SALT) 32*, volume 32, forthcoming.

[7] Moshe E Bar-Lev and Danny Fox. Universal free choice and innocent inclusion. In *Semantics and Linguistic Theory (SALT) 27*, volume 27, pages 95–115, 2017.

[8] Moshe E Bar-Lev and Daniel Margulis. Hebrew *kol*: a universal quantifier as an undercover existential. In *Proceedings of Sinn und Bedeutung (SuB) 18*, volume 18, pages 60–76, 2014.

[9] Gennaro Chierchia. *Logic in Grammar*. Oxford University Press, Oxford, 2013.

[10] Barbara H Partee. Opacity and scope. In Milton Karl Munitz and Peter K. Unger, editors, *Semantics and Philosophy*, pages 81–101. New York University Press, 1974.

[11] Florian Schwarz. Intensional transitive verbs: *I owe you a horse*. *The Wiley Blackwell Companion to Semantics*, pages 1–33, 2020.

Dynamic Relationism about Belief*

Dilip Ninan

Tufts University, Medford, MA, USA
dilip.ninan@tufts.edu

Abstract

This essay presents a novel account of intentional identity, the phenomenon illustrated by Geach's 'Hob-Nob' sentence. On the account defended here, such sentences express dyadic belief facts that are not in general reducible to monadic belief facts. We show how these non-reductive truth-conditions are predicted by a version of dynamic semantics which allows sentences to update certain kinds of accessibility relations.

1 Introduction

The problem of intentional identity concerns Geach's infamous 'Hob-Nob' sentence:

(1) [There are no witches, but] Hob believes that a witch blighted Bob's mare, and Nob believes that she killed Cob's sow. [4]

This example raises two questions. First, what are the truth-conditions of (1)? Second, how are those truth-conditions compositionally generated from the meanings of its parts? We begin with a brief discussion of extant answers to the first question, and then suggest taking a rather different approach altogether. On the approach taken here, facts of the form *a believes that* ϕ *and b believes that* ψ are not reducible to facts of the form *c believes that* χ. After briefly explaining and motivating this 'relationist' view, we present a version of dynamic semantics that yields our proposed truth-conditions for (1).

2 Descriptivism

On the intended reading of (1), the pronoun *she* in the second conjunct of (1) is not a deictic pronoun; it is in some sense anaphoric on the indefinite noun phrase *a witch* that occurs in first conjunct. And the intended reading is not one that requires there to be any witches—hence the bracketed conjunct *there are no witches*. But what *does* (1) require for its truth? What are the truth-conditions of this sentence?

It is natural to wonder whether the intended reading of (1) can be paraphrased by replacing the pronoun *she* in the second conjunct with a definite description constructed from the linguistic material found in the first (unbracketed) conjunct. Perhaps (1) is equivalent to (2) or to (3):

(2) Hob believes that a witch blighted Bob's mare and Nob believes that *the witch that blighted Bob's mare* killed Cob's sow.

(3) Hob believes that a witch blighted Bob's mare and Nob believes that *the witch that Hob believes blighted Bob's mare* killed Cob's sow.

*For comments on an earlier version, thanks to an audience at PhLiP 7 (Tarrytown, NY), and to Simon Charlow, Josh Dever, Andy Egan, Matt Mandelkern, and Sarah Moss.

(All of the belief ascriptions here should be read *de dicto*.) While each of these sentences might capture a reading of (1), Geach argued in his original article that there is a reading of (1) on which it is equivalent to neither of these. For it seems that there are situations in which (1) is true while both (2) and (3) are false. Here is an example, adapted from [2]:

> **Newspaper Case**
>
> A number of animals in Gotham Village have recently died quite unexpectedly. Rumors have begun to circulate that these unfortunate events are due to the machinations of a witch. The local newspaper, the Gotham Star, has picked up on these rumors and reported that the witch has been attacking animals and destroying crops. In reality, there is no such person: the animals in question all died of natural causes, the crops withered from drought. Hob and Nob both read the Gotham Star and both believe the article about the witch. Hob thinks that the witch must have blighted Bob's mare, which fell ill recently, while Nob thinks that the witch killed Cob's sow. But Nob is unaware of Hob's and Bob's existence, and so has no beliefs about Hob or Bob at all.

It is clear from this last detail of the story, that neither (2) nor (3) is true in this scenario. Since Nob has no beliefs about either Hob or Bob, neither (2) nor (3) can be true. But it is widely thought that (1) has a reading on which it is true in this scenario.[1] If that is correct, then neither (2) nor (3) captures the intended reading of (1).

Examples like the Newspaper Case have been taken to show that, on the intended reading, (1) requires that Hob's belief and Nob's belief have a *common causal source* [5, 8].[2] For note that, in the Newspaper Case, Hob's belief and Nob's belief are both partially caused by the articles in the Gotham Star, or by the rumors circulating in town. This observation can be used to motivate an alternative 'descriptivist' story. On this approach, (1) is equivalent to something like:

(4) Hob believes that a witch blighted Bob's mare and Nob believes that *the witch described by the actual common causal source of Hob's belief and Nob's belief* killed Cob's sow. [8, 298]

The inclusion of the adjective *actual* is intended to make (4) (and thus (1)) equivalent to:

(5) Hob believes that a witch blighted Bob's mare and the common causal source S of Hob's belief and Nob's belief is such that Nob believes that *the witch described by* S killed Cob's sow.

Note that if (5) is true, then *the common causal source of Hob's belief and Nob's belief* is a non-empty definite description. Since (1) is equivalent to (5) on this proposal, the account predicts that the truth of (1) entails that Hob's belief and Nob's belief have a common causal source. Furthermore, this proposal avoids the problem facing the previous descriptivist approaches, since its proposed truth-condition doesn't require that Nob is aware of Hob or Bob, only that he is aware of the causal source of his belief.

The principal difficulty with this approach is that while (1) might require that Hob's belief and Nob's belief *have* a common causal source, it would not appear to require that *Nob believe*

[1] Though see [7] and [1] for some doubts about this.

[2] Here and in what follows, by *Hob's belief* I mean Hob's belief that a witch blighted Bob's mare, and by *Nob's belief* I mean Nob's belief that someone killed Cob's sow.

anything about this source. For example, while it is natural to assume that, in the Newspaper Case, Nob has a *de dicto* belief to the effect that the witch described in the Gotham Star article killed Cob's sow, this is not obligatory. Imagine, for example, that Nob reads the article in the Gotham Star, forms the belief that the witch killed Cob's sow, and then proceeds to forget how he formed this belief. Maybe he later comes to believe that he learned about the witch from his friend Janice, or maybe he simply forms no new beliefs about the source of his witch-beliefs.[3] The article in the Gotham Star is the causal source S of Hob's belief and of Nob's belief, but since Nob has forgotten all about that article, he has no beliefs about S. So (5) is false is this version of the Newspaper Case. Nevertheless, it seems that (1) is still true, which suggests that (1) is not equivalent to (5) after all.

3 Dynamic relationism

Within the framework of possible worlds semantics, a sentence of the form *a believes that* ϕ is said to be true iff in every world w compatible with what a believes, ϕ. Framed in this way, our problem becomes the problem of saying what it is for a world w to be compatible with what Hob believes and what it is for a world w' to be compatible with what Nob believes, given that (1) is true. The first of these questions has a natural answer: w should contain a witch who blighted Bob's mare. The trouble comes with saying what w', an arbitrary world compatible with what Nob believes, should be like. World w' should contain someone y who killed Cob's sow, but who in w' is y? The various answers to this we've considered (e.g. *y is the witch who blighted Bob's mare in w'*) have all proved unsatisfactory.

While there might be other answers to this question worth considering,[4] I suggest that we change course and ask a different question altogether. Suppose we instead ask: what is it for a *pair* of worlds (w, w') to be compatible with what the pair (Hob, Nob) believe? This question turns out to have a comparatively natural answer: (w, w') should be compatible with what (Hob, Nob) believe only if there is an x such that x is a witch in w, x blighted Bob's mare in w, and x killed Cob's sow in w'. The view taken here is (to a first approximation) that (1) is true iff every pair (w, w') compatible with what (Hob, Nob) believe meets this condition.

This view will be developed below, but let me pause here to explain what I intend by calling this a *relationist* theory. Let's say that an agent *a has precisely the same monadic beliefs* in w as they have in w' iff for all ϕ, a believes that ϕ in w iff a believes that ϕ in w'.[5] And let's say that agents a and b *stand in precisely the same dyadic belief relations* in w as they do in w' iff for any ϕ and ψ, a believes that ϕ and b believes that ψ in w iff a believes that ϕ and b believes that ψ in w'. Then *reductionism about dyadic belief* is the view that for any worlds w and w' and agents a and b, if a has precisely the same monadic beliefs in w as they have in w', and b has precisely the same monadic beliefs in w as they have in w', then a and b stand in precisely the same dyadic belief relations in w as they do in w'. *Relationism about dyadic belief* is the negation of reductionism.[6]

It might seem as if reductionism is inescapable, but the approach to (1) taken above suggests how relationism could be true. To see what I mean, let a and b be fixed but arbitrary agents. Let $Dox^w_{a,b}$ be the set of pairs of worlds compatible with what (a, b) believe in w, let $Dox^w_a :=$

[3] We often forget how we formed certain beliefs. I believe that a moose once wandered into a school in Saskatoon, but I forget how I formed this belief. (Did I read it in the paper? Did a friend tell me?)

[4] See, for example, [9], [10], and [5].

[5] We state this definition, and the views to follow, in a species of mathematical English that permits quantification into sentence position.

[6] My use of the term *relationism* is inspired by [3], though my version of relationism is rather different from Fine's.

$\{v : (v, v') \in Dox^w_{a,b}$, for some $v'\}$, and let $Dox^w_b := \{v' : (v, v') \in Dox_{a,b}$, for some $v\}$. Assume that for any $c \in \{a, b\}$, c has precisely the same monadic beliefs in w as they have in w' iff $Dox^w_c = Dox^{w'}_c$. And assume that a and b stand in precisely the same dyadic belief relations in w as they do in w' iff $Dox^w_{a,b} = Dox^{w'}_{a,b}$. Given certain assumptions about the space of worlds, we can show that there are worlds w and w' such that $Dox^w_a = Dox^{w'}_a$, $Dox^w_b = Dox^{w'}_b$, but $Dox^w_{a,b} \neq Dox^{w'}_{a,b}$. In that case, a and b will each have precisely the same monadic beliefs in w as they have in w', but they will not stand in precisely the same dyadic belief relations in w as they do in w'. Relationism will be true, reductionism false.

To see how this would work, suppose there are worlds w and w' such that:

$Dox^w_{a,b} = \{(v, v') : \exists x$ (x blighted Bob's mare in v and x killed Cob's sow in v')$\}$.

$Dox^{w'}_{a,b} = \{(v, v') : \exists x$ (x blighted Bob's mare in v) and $\exists y$ (y killed Cob's sow in v')$\}$.

It seems that $Dox^w_{a,b} \neq Dox^{w'}_{a,b}$. For suppose that in v, x alone blighted Bob's mare, and that in v', y alone killed Cob's sow, where $y \neq x$. Then (v, v') will be in $Dox^{w'}_{a,b}$, but not in $Dox^w_{a,b}$ Thus, a and b will not stand in precisely the same dyadic belief relations in w as they do in w'. But we also have that $Dox^w_a = Dox^{w'}_a$ and $Dox^w_b = Dox^{w'}_b$, which means that a and b each have precisely the same monadic beliefs in w as they have in w'.[7] Thus, we have a difference in the relevant dyadic belief facts despite no difference in the relevant monadic belief facts.

The Newspaper Case can be used to provide intuitive motivation for relationism. For we might take that case to show that (1) is true iff: (i) Hob believes that a witch blighted Bob's mare, (ii) Nob believes that someone killed Cob's sow, and (iii) Hob's belief and Nob's belief have a common causal source (of the appropriate kind). That might be too simplistic, but if *something* like this is right, then it should not be too difficult to construct a pair of scenarios w and w' such that Hob has the precisely the same monadic beliefs in w as he has in w', Nob has the precisely the same monadic beliefs in w as he has in w', but (1) is true in w but not in w'. For suppose that (i) and (ii) both obtain in w and w', while (iii) obtains in w but not in w'. Then it is plausible to think that (1) will be true in w, false in w'. But all that appears to be compatible with Hob and Nob each having precisely the same monadic beliefs in w as they have in w'. For the key difference between w and w' concerns the *causal origins* of Hob's belief or of Nob's belief. And it's plausible to think the causal origins of one of those beliefs could differ across w and w' even if Hob and Nob each have precisely the same monadic beliefs in w as they have in w'. Whether or not Hob's belief and Nob's belief have a common causal source is a fact *external* to their monadic beliefs; thus, the presence or absence of a common cause may vary even as their monadic beliefs are held fixed.

While there is more to be said about these foundational matters, we turn now to develop the relationist approach in more detail. In order to generalize the approach, first note that there is nothing special about pairs of believers:

(6) Hob believes that a witch blighted Bob's mare, Nob believes that she killed Cob's sow, and Joe believes that she stole Janice's tractor.

[7] To see that $Dox^w_a \subseteq Dox^{w'}_a$, suppose $v \in Dox^w_a$. Then there is a v' and an x such that x blighted Bob's mare in v and x killed Cob's sow in v'. But then there is there is a z that blighted Bob's mare in v and there is a y that killed Cob's sow in v', for x is such a z and such a y. So $(v, v') \in Dox^{w'}_{a,b}$, which means $v \in Dox^{w'}_a$. To see that $Dox^{w'}_a \subseteq Dox^w_a$, suppose $v \in Dox^{w'}_a$. So there is a v' such that there is an x that blighted Bob's mare in v and there is a y that killed Cob's sow in v'. Let u be a world in which x killed Cob's sow (we assume, plausibly, that there is such a world). So x blighted Bob's mare in v and x killed Cob's sow in u. So $(v, u) \in Dox^w_{a,b}$, which means $v \in Dox^w_a$, as desired. The argument that $Dox^w_b = Dox^{w'}_b$ is similar.

That suggests that we should speak of an n-ary sequence of worlds $(w_1, ..., w_n)$'s being compatible with what an n-ary sequence of individuals $(a_1, ..., a_n)$ believe. But since nothing in the present phenomenon mandates the order built into these sequences, I propose that we use *functions* from individuals to worlds. Given a non-empty set W of worlds and a non-empty set of agents A, let $\mathsf{f} \in W^A$ be a *doxastic alternative for* A.[8] Then if {Hob, Nob} $\subseteq A$, I propose that (1) is true at a world w iff for all doxastic alternatives $\mathsf{f} \in W^A$, there is an x such that x is a witch who blighted Bob's mare in f(Hob) and x killed Cob's sow in f(Nob).

We turn now to the task of constructing a compositional semantics that generates these truth conditions. Our theory is a variant of Dynamic Predicate Logic [6] that allows the indefinite *a witch* in the first conjunct of (1) to control the interpretation of the pronoun *her* in the second conjunct. This is achieved by allowing a clause of the from $\mathcal{B}_b\exists x\phi$ to update a certain type of relation that we call an *accessibility relation*. These accessiblity relations hold between points that include variable assignments.

Given a non-empty set of agents A, we assume a language $\mathcal{L}_A$. (In the intended application, A would be the set of all actual and possible agents.) The vocabulary of this language consists of n-ary relation symbols, variables, $\neg$, $\wedge$, $\exists x$, and, for each $a \in A$, a belief operator $\mathcal{B}_a$. The definition of the formulas of $\mathcal{L}_A$ can be gleaned from the recursive semantics below. We translate (1) into this language as $(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$.

Definition 1. A *model* for $\mathcal{L}_A$ is a tuple $M = (W, D, B, I)$ where:

(1) W is a non-empty set of worlds,

(2) D is a non-empty set of individuals,

(3) B is a relation between worlds $w \in W$ and elements $\mathsf{f} \in W^A$ (i.e. $B \subseteq W \times W^A$), and

(4) I is a function which maps an n-ary relation symbol and a world to a subset of D^n.

Regarding (3): the idea is that, in an intended model, $wB\mathsf{f}$ iff f is compatible with what the agents in A believe in w.

The definitions that follow are given relative to a fixed model $M = (W, D, B, I)$ for a fixed language $\mathcal{L}_A$.

A variable assignment is a function from the variables of $\mathcal{L}_A$ into D. We use G to denote the set of variable assignments. If g and h are variable assignments and x a variable, we say that h is an *x-variant* of g, $h[x]g$, iff for all variables y other than x, $h(y) = g(y)$.

Definition 2. A binary relation $R \subseteq (W \times G) \times (W^A \times G)$ is an *accessibility relation* iff: (i) $(w, g)R(\mathsf{f}, g')$ only if $wB\mathsf{f}$, and (ii) if $wB\mathsf{f}$, then for any g, there is a g' s.t. $(w, g)R(\mathsf{f}, g')$. Let $R(w, g) := \{(\mathsf{f}, g') : (w, g)R(\mathsf{f}, g')\}$

Definition 3. We define the semantic value of a formula ϕ relative to: a doxastic alternative $\mathsf{f} \in W^A$, an agent $a \in A$, a pair of variable assignments $g, g' \in G$, and a pair of accessibility relations R, R'.

(1) $[\![Px_1, ..., x_n]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $g = g'$, $R = R'$, and $(g(x_1), ..., g(x_n)) \in I(P, \mathsf{f}a)$

(2) $[\![\neg\phi]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $g = g'$, $R = R'$, and there is no h and Q such that $[\![\phi]\!]^{\mathsf{f},a,g,h,R,Q} = 1$

(3) $[\![\phi \wedge \psi]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff there is an h and a Q such that $[\![\phi]\!]^{\mathsf{f},a,g,h,R,Q} = 1$ and $[\![\psi]\!]^{\mathsf{f},a,h,g',Q,R'} = 1$

[8] Where W^A is the set of all functions $\mathsf{f} : A \to W$.

(4) $[\![\exists x\phi]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $R = R'$ and there is an h such that $h[x]g$ and $[\![\phi]\!]^{\mathsf{f},a,h,g',R,R'} = 1$[9]

(5) $[\![\mathcal{B}_b\phi]\!]^{\mathsf{f},a,g,g',R,R'} = 1$ iff $g = g'$ and:

(a) for all $(\mathsf{f}', h) \in R(\mathsf{f}a, g)$, there is an h' and a Q such that $[\![\phi]\!]^{\mathsf{f}',b,h,h',R,Q} = 1$, and

(b) for all $(\mathsf{f}', h) \in R'(\mathsf{f}a, g)$, there is a h' such that $(\mathsf{f}', h') \in R(\mathsf{f}a, g)$ and for some Q, $[\![\phi]\!]^{\mathsf{f}',b,h',h,R,Q} = 1$

Given a world w, let f^w be the element of W^A such that $\mathsf{f}^w a = w$, for all $a \in A$. And let R_{max} be the accessibility relation R such that for any world w and function $\mathsf{f} \in W^A$, if $wB\mathsf{f}$, then $(w, g)R(\mathsf{f}, g')$, for any variable assignments g and g'. Then we can define *truth at a world* as follows:

Definition 4. A sentence ϕ is *true at a world* w iff for any g and any $a \in A$, there is a (g', R') such that $[\![\phi]\!]^{\mathsf{f}^w,a,g,g',R,R'} = 1$, where $R = R_{max}$.

This view predicts that (1) has our proposed truth-conditions:

Proposition 1. *$(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$ is true at a world w iff for all $\mathsf{f} \in W^A$, if $wB\mathsf{f}$, then there is an $o \in D$ such that $o \in I(F, \mathsf{f}b)$ and $o \in I(G, \mathsf{f}c)$.*

Proof-sketch of the left-to-right direction. Suppose $(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)$ is true at a world w. From the definition of truth at a world, it follows that there is a g' and and R' such that:

$$[\![(\mathcal{B}_b\exists xFx \wedge \mathcal{B}_cGx)]\!]^{\mathsf{f}^w,a,g,g',R,R'} = 1.$$

(Here a and g are arbitrary and $R = R_{max}$.) It follows from this and from the clause for $\wedge$ that there is a k and a Q such that:

(i) $[\![\mathcal{B}_b\exists xFx]\!]^{\mathsf{f}^w,a,g,k,R,Q} = 1$, and

(ii) $[\![\mathcal{B}_cGx]\!]^{\mathsf{f}^w,a,k,g',Q,R'} = 1$.

From (i) and the clauses for $\mathcal{B}$ (5b), $\exists x$, and atomics, it follows that $g = k$ and that:

(A) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$: $h(x) \in I(F, \mathsf{f}b)$.

Since $g = k$, (ii) is equivalent to $[\![\mathcal{B}_cGx]\!]^{\mathsf{f}^w,a,g,g',Q,R'} = 1$. And given the clauses for $\mathcal{B}$ (5a) and atomics, this implies:

(B) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g) : h(x) \in I(G, \mathsf{f}c)$.

Claims (A) and (B) together imply:

(C) for all $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$: $h(x) \in I(F, \mathsf{f}b)$ and $h(x) \in I(G, \mathsf{f}c)$.

Now suppose $wB\mathsf{f}$. Note that $w = \mathsf{f}^w a$, so we are supposing that $\mathsf{f}^w aB\mathsf{f}$. Since Q is an accessibility relation, there is an h such that $(\mathsf{f}, h) \in Q(\mathsf{f}^w a, g)$. So let (f, h) be any element of $Q(\mathsf{f}^w a, g)$. So from (C) we have that $h(x) \in I(F, \mathsf{f}b)$ and $h(x) \in I(G, \mathsf{f}c)$. So there is an $o \in D$ such that $o \in I(F, \mathsf{f}b)$ and $o \in I(G, \mathsf{f}c)$, for $h(x)$ is such an o. □

[9] A slightly more complex clause is needed here to handle *de re* belief attributions, but I set this issue aside for simplicity.

References

[1] David Braun. Hob, Nob, and mythical witches. In Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein, editors, *Reference and Referring*, pages 149–87. MIT Press, Cambridge, MA, 2012.

[2] Walter Edelberg. A new puzzle about intentional identity. *Journal of Philosophical Logic*, 15:1–25, 1986.

[3] Kit Fine. *Semantic Relationism*. Blackwell Publishing Ltd, Oxford, 2009.

[4] Peter T Geach. Intentional identity. *The Journal of Philosophy*, 64(20):627–632, 1967.

[5] Ephraim Glick. A modal approach to intentional identity. *Noûs*, 46(3):386–399, 2012.

[6] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14:39–100, 1991.

[7] Jeffrey C King. Intentional identity generalized. *Journal of Philosophical Logic*, pages 61–93, 1993.

[8] William Lanier. Intentional identity and descriptions. *Philosophical Studies*, 170(2):289–302, 2014.

[9] Nathan Salmon. Mythical objects. In J. Campbell, M. O'Rourke, and D. Shier, editors, *Meaning and Truth*. Proceedings of the Eastern Washington University and the University of Idaho Inland Northwest Philosophy Conference on Meaning, Seven Bridges Press, 2002.

[10] Robert van Rooij. *Attitudes and Changing Contexts*. Springer, Dordrecht, 2006.

Non-past Directives in Japanese*

Yuya Noguchi

University of Connecticut
yuya.noguchi@uconn.edu

Abstract

This paper concerns non-past directives (NPDs) in Japanese, non-past declarative sentences that serve as directive speech acts like imperatives. I observe that NPDs can be felicitous in a more restricted range of contexts than imperatives. To capture this restriction, building on Kaufmann's (2012) modal approach to imperatives, I propose that NPDs contain a modal that is the same as the one in imperatives except that it triggers a presupposition that the speaker considers the behavior of the addressee(s) to be irrational. This proposal will be supported by the "Hey, wait a minute." test. The results of this paper tentatively support the modal approach to the semantics of imperatives.

1 Introduction

In Japanese, declarative sentences with a verb followed by a non-past morpheme -(*r*)*u* (a non-past verb) usually describe a present state or future event, as exemplified in (1). [1]

(1) Ken-wa {mainichi / asita} hasir-u.
Ken-TOP every.day tomorrow run-NP
'Ken runs {every day / tomorrow}.'

However, they can serve as directive sentences, usually accompanied by phonological emphasis (e.g. Arita 2015, Noguchi 2016, Ihara & Noguchi 2018, Ihara 2020, 2021). For example, (2), where the capitals represent phonological emphasis, is interpreted as an order, as is the (standard) imperative in (3).

(2) HASIR-U.
run-NP
'Run.'

*I would like to thank Jon Gajewski, Magdalena Kaufmann and Stefan Kaufmann for helpful discussion and insightful comments. I am also grateful to the anonymous reviewers of Amsterdam Colloquium 2022 for helpful comments. All remaining errors are, of course, my own.

[1] The following abbreviations are used: ACC = accusative, GEN = genitive, IMP = imperative morpheme, NOM = nominative, NP = non-past morpheme, POL = politeness, Q = question marker, SFP = sentence final particle, TOP = topic particle

(3) Hasir-e.
run-IMP
'Run.'

This paper calls directive sentences like (2) *non-past directives* (*NPDs*). To the extent that they have been discussed in the literature, it is assumed that NPDs are generally interchangeable with corresponding imperatives.[2] However, this paper shows a new observation that NPDs are contextually more restricted than imperatives. Building on this observation, this paper proposes a semantic analysis of NPDs.

This paper is organized as follows: Section 2 illustrates that NPDs are felicitous only in a limited range of contexts, compared with imperatives. Section 3 proposes a semantic analysis of NPDs based on Kaufmann's (2012) modal approach to imperatives, and provides support for the proposal. Section 4 concludes the paper.

2 Contextual restrictions

This section illustrates the contextually restricted use of NPDs. Observe first that both the NPD (2) and the imperative (3) are felicitous in the contexts (4) and (5).

(4) [It is a well-known rule in this PE course that students must run around the grounds when a class starts. In one class, the students somehow feel too lazy to run, keeping chatting. When the class starts, the teacher is surprised to find that they have not started to run. He says to them:]
the NPD (2): ✓ / the imperative (3): ✓

(5) [The teacher mistakenly believes that the new first-year students know the rule of his PE course that students must run around the grounds when a class starts. So, when the first class starts, the teacher is surprised to find that they have not started to run. He says to them:]
the NPD (2): ✓ / the imperative (3): ✓

In the following contexts, however, only the imperative (3) is felicitous:[3]

(6) [The teacher is planning to make a rule for his new PE course that the students run around the grounds as a first exercise. In the first class, he says to them:]
the NPD (2): # / the imperative (3): ✓

(7) [It is a well-known rule in this PE course that students must run around the grounds when a class starts. In one class, the teacher arrives a little early and so the students have not yet started to run. He expects them to start running soon since they have followed the rule in every class. As a reminder, he says to them:]
the NPD (2): # / the imperative (3): ✓(with the sentence final particle *yo*)

[2] Ihara & Noguchi (2018) and Ihara (2020, 2021) observe some differences between imperatives and NPDs, such as the availability of so-called weak readings like permissions. As these do not concern the core directive uses that I am concerned with in this paper, I will leave it to future work to consider how my proposal relates to their observations.

[3] I leave open why the sentence final particle *yo* is required for the imperative (3) in the context (7) (see Davis 2009, 2011 and Oshima 2014 for relevant discussion and further references).

To identify the crucial difference between (4-5) and (6-7), notice first that (4) and (5) are common in that the students seem to the teacher to be "lazy", though they differ in whether his assumption about their laziness is correct or not. More precisely, this can be described as in (8).

(8) a. The speaker (i.e. the teacher) believes that it is obvious to the addressee(s) (i.e. the students) that they should realize the prejacent *p* (i.e. *you run*) in the current situation.
b. It seems to the speaker that the addressee(s) will not realize *p* in the current situation.

Crucially, (6-7) differ from (4-5) in whether (8) holds. In (6), (8a) is not the case (and accordingly, (8b) is not either); the teacher knows that the students do not yet know the planned rule, thus not believing that it is obvious for them that they should run. In (7), (8a) holds but (8b) does not; the teacher assumes that it is obvious to the students that they should run, and he has no reason to think that they will not do so, given that they have been following the rule on all previous occasions. The contrasts suggest that NPDs are felicitous only in contexts characterized by (8a) and (8b), unlike imperatives. The next section aims to propose a semantic analysis of NPDs that can capture this contextual restriction.

3 Proposal

3.1 Modal approach to imperatives

For the proposal of this paper, I adopt the modal approach to the semantics of imperatives (e.g. Kaufmann 2012). According to this approach, imperatives involve the imperative modal operator *IMP*, which takes a propositional prejacent *p* as in (9a). This paper assumes that utterance contexts are characterized by what is common ground between the interlocutors *CG* (Stalnaker 1978), the world *w* and time *t* at which they take place, and that they determine a salient modal base *f* and an ordering source *g*. With this, putting aside details irrelevant for the following discussion, I assume that IMP is interpreted as in (9b), based on the Kratzer-style analysis of modals (e.g. Kratzer 1991, 2012).[4]

(9) a. [IMP [*p*]]
b. $[\![(9a)]\!]^{c} = \forall w' \in BEST(CG_c, g_c, t_c, w_c)[\exists t'[P(t', w') \,\&\, t_c<t']]$
presupposes, among others:
p resolves a salient decision problem of the addressee(s) (consisting of future courses of their actions), such that the speaker and the addressee(s) consider the ordering source *g* the relevant criterial for solving the decision problem. (**Ordering Source Restriction**: OSR)

$BEST(CG_c, g_c, t_c, w_c)$ in (9b) denotes a set of worlds which are compatible with the common ground CG_c (as a modal base) and best according to the contextually salient ordering source g_c at the utterance time t_c in the actual world w_c. The at-issue meaning of (9b) is then that *p* is true in all the worlds of $BEST(CG_c, g_c, t_c, w_c)$. The at-issue meaning of IMP is thus the same as necessity modals like *should* and *must*. It differs from those modals, however, in that it triggers presuppositions that make imperatives non-assertoric, or performative. The only presupposition

[4] For expository purposes, I assume that the prejacent of an imperative (and an NPD) involves a tense element that posits the event denoted by the prejacent in a temporal location which follows the context time t_c, as shown in (9b). Kaufmann (2012) argues that the temporal interpretation of imperatives is represented in the presuppositional component.

that is relevant to the following discussion is *Ordering Source Restriction* (*OSR*); see Kaufmann (2012) for other presuppositions.

With this approach to imperatives, I assume that the Japanese imperative (3), repeated below, is analyzed to have the structure in (10a), and (10a) is interpreted as in (10b).[5,6]

(3) Hasir-e.
run-IMP
'Run.'

(10) a. [IMP [you run]]
b. $[\![(10a)]\!]^{c} = \forall w' \in BEST(CG_c, g_c, t_c, w_c)[\exists t'[\text{you.run}(t', w') \ \& \ t_c < t']]$
presupposes, among others:
p (i.e. *you run*) resolves a salient decision problem of the addressee(s) (consisting of future courses of their actions), such that the speaker and the addressee(s) consider the ordering source *g* the relevant criterial for solving the decision problem. (**Ordering Source Restriction**: OSR)

3.2 Special imperative modal

Building on the modal approach to imperatives illustrated above, I now propose that NPDs involve a special type of IMP, *S*(*pecial*)-*IMP*. More specifically, I argue that the NPD in (2), repeated below, has the structure in (11a), and (11a) is interpreted as in (11b).

(2) HASIR-U.
run-NP
'Run.'

(11) a. [S-IMP [you run]]
b. $[\![(11a)]\!]^{c} = \forall w' \in BEST(CG_c, g_c, t_c, w_c)[\exists t'[\text{you.run}(t', w') \ \& \ t_c < t']]$
presupposes, among others:
(i) *p* (i.e. *you run*) resolves a salient decision problem of the addressee(s) (consisting of future courses of their actions), such that the speaker and the addressee(s) consider the ordering source *g* the relevant criterial for solving the decision problem. (**Ordering Source Restriction**: OSR)
(ii) The addressee(s) is(/are) behaving irrationally with respect to their decision problem (**Irrational Behavior by Addressee**: IBA).

S-IMP is the same as IMP except that it triggers an additional presupposition *Irrational Behavior by Addressee* (*IBA*). The notion "behaving irrationally" draws on Kaufmann & Kaufmann's (2012) proposal that presupposing OSR entails *Rational Choice* (12) (where $\Box^{f,g}q$ stands for a proposition where the prejacent *q* is modalized by a necessity modal like *should* and *must* with respect to the modal base *f* and the ordering source *g*). Note in particular that (12) characterizes a *rational* hearer.

(12) A rational hearer who believes $\Box^{f,g}q$ such that *q* serves as a solution to the salient decision problem will aim to bring about *q*. (*Rational Choice*)
(cf. Kaufmann & Kaufmann 2012: 219)

[5] See Ihara & Noguchi (2018) and Ihara (2020, 2021) for the modal approach to imperatives in Japanese.
[6] I assume that imperatives (and NPDs) involve a covert second person pronoun as their subject, as in (10).

Encoding IBA in S-IMP is motivated by the characteristics of the contexts (4-5), where the NPD (2) is felicitous. In these contexts, the teacher (correctly or wrongly) believes that the students know that they should run, based on his assumption that they know the rule. In addition, the teacher should presuppose OSR in (4-5), given that imperatives, which presuppose OSR (see (9/10b)), are felicitous as well in those contexts. Rational Choice (12) should thus be entailed in (4-5). It then follows from these characteristics (i.e. (i) the teacher's assumption that the students know that they should run, (ii) the entailment of Rational Choice) that in (4-5), the teacher assumes that the students will aim to run, as long as they are *rational*. Actually, however, it seems to him that they will not run. The teacher thus concludes that they are *irrational* and presupposes this, namely IBA, when using the NPD.

Let us now consider how the current proposal can account for the felicity difference of the NPD (2) between (4-5) and (6-7). On the one hand, it is observed in Section 2 that (4-5) and (6) differ in whether the teacher believes that it is obvious to the students that they should run (see (8a)); he believes so in (4-5) because of the teacher's assumption that the students know the rule, while he does not in (6) because the teacher has not told the students about the planned rule. Note here that if the speaker is to determine whether the addressee(s) is(/are) rational based on Rational Choice (12), (s)he needs to assume that they believe $\Box^{f,g}q$ (and to find out whether they are likely to bring about q). Therefore, in (6) there is no reason for the teacher to assume that the students are irrational, preventing him from presupposing IBA. On the other hand, as observed in Section 2, while (7) is similar to (4-5) in that the teacher believes that it is obvious for the students that they should run, they differ regarding the behavior of the students (see (8b)); in (4-5) it seems to the teacher that the students will not run, but in (7) he has no reason to think that they will not do so, given their obedience to the rule. Hence, in (7), the attitude of the students does not seem to the teacher to be irrational in the sense of (12) and accordingly he cannot presuppose IBA. The infelicity of NPDs in (6-7) can thus be accounted for by presupposition failure with respect to IBA. Thus, the current proposal captures the contextual restriction of NPDs.

3.3 Support

A piece of evidence for the proposal that NPDs involve a special imperative modal that triggers IBA can be provided by the "Hey, wait a minute." test.[7] Von Fintel (2004) argues that when one states a complaint after saying "Hey, wait a minute.", that complaint is regarding what the speaker presupposes, rather than what (s)he asserts. In (13), for example, A presupposes that someone proved Goldbach's Conjecture when he makes the utterance, as suggested by the subject *the mathematician who proved Goldbach's Conjecture*. After saying "Hey, wait a minute.", B can felicitously complain about A's making that presupposition. On the other hand, B's statement would be infelicitous when she said, after "Hey, wait a minute.", that she did not know about the asserted content.

(13) A: The mathematician who proved Goldbach's Conjecture is a woman.
B: Hey, wait a minute. I had no idea that someone proved Goldbach's Conjecture.
B': # Hey, wait a minute. I had no idea that that was a woman.
(von Fintel 2004: 271)

This test applies in Japanese as well; translating (13) into Japanese results the same, as (14) shows.

(14) A: Goorudobahha-no yosoo-o shoomeesita suugakusya-wa zyosee nanda tte.
Goldbach-GEN conjecture-ACC proved mathematician-TOP woman is SFP
'The mathematician who proved Goldbach's Conjecture is a woman.'

[7] I thank Yusuke Yagi for suggesting applying the "Hey, wait a minute." test to NPDs.

B: E, chotto matte. Goorudobahha-no yosoo-o shoomeesita hito-ga
hey little wait Goldbach-GEN conjecture-ACC proved person-NOM
iru nante siranakatta yo.
there.is C not.knew SFP
'Hey, wait a minute. I didn't know that (there is) someone (who) proved Goldbach's Conjecture.'

B': # E, chotto matte. Sono hito-ga zyosee da nante siranakatta yo.
hey little wait that person-NOM woman is C not.knew SFP
'Hey, wait a minute. I didn't know that that person was a woman.'

With this background, let us observe the data (15), where one student makes a complaint after "Hey, wait a minute.", following the teacher's utterance of the NPD.

(15) [The teacher mistakenly believes that the new first-year students know the rule of his PE course that students must run around the grounds when a class starts. So, when the first class starts, the teacher is surprised to find that they have not started to run.]
Teacher: HASHIR-U.
run-NP
'Run.'
Student: E, chotto mat-tekudasai. Nani-o suru beki ka wakaranakatta kara,
hey little wait-IMP.POL what-ACC do should Q not.knew because
hasiranakatta n desu yo.
not.ran C is.POL SFP
'Hey, wait a minute. We didn't run because we didn't know what to do.'

Note that the context in (15) is the same as (5). In that context, as discussed in Section 3.2, the teacher wrongly believes that the students know the rule, and thus considers their not running to be irrational (satisfying IBA). Crucially, in (15), after saying "Hey, wait a minute.", the student felicitously complains about the teacher's wrong belief by explaining the reason why they did not run, stating that that was because they did not know what to do (not knowing the rule). The felicity of the complaint thus suggests that NPDs should involve IBA in their presuppositional component.

This suggestion is further corroborated by the data in (16), which is the same as (15) except that the teacher utters the corresponding imperative, not the NPD.

(16) [In the same context as (15):]
Teacher: Hashir-e.
run-IMP
'Run.'
Student: # E, chotto mat-tekudasai. Nani-o suru beki ka wakaranakatta kara,
hey little wait-IMP.POL what-ACC do should Q not.knew because
hasiranakatta n desu yo.
not.ran C is.POL SFP
'Hey, wait a minute. We didn't run because we didn't know what to do.'

(16) crucially shows that the student cannot make the same complaint as in (15) if the preceding utterance by the teacher is an imperative; the students would accept the imperative in (16) as a mere order by the teacher that the students should run. The difference between (15) and (16) thus indicates that IBA is triggered by NPDs, but not by imperatives.

4 Conclusion

To recap, this paper has provided a hitherto unacknowledged fact that in contrast with imperatives, NPDs are felicitous only in a limited range of contexts, where the speaker assumes that the addressee(s) is(/are) "lazy". Based on this observation, I have argued, adopting the modal-based analysis of imperatives (e.g. Kaufmann 2012), that NPDs involve the special imperative modal S-IMP, which triggers a presupposition that the speaker considers the behavior of the addressee(s) irrational, namely IBA. I have further shown that this proposal is supported by the "Hey, wait a minute." test.

Note finally that the contextual restriction of NPDs is not observed in declaratives used as directive in other languages (e.g. *You will run!* in English). This difference supports the idea that the directive construal of NPDs results from a modal operator which can be parameterized, rather than from pragmatic principles. Given the overall similarity between NPDs and imperatives, the proposal may be considered tentative support of the modal approach to, or a "strong" theory of, imperatives overall (see von Fintel & Iatridou 2017 for discussion; Portner 2004 for an alternative "minimal" theory).

References

Arita, Setsuko. 2015. "Sasisematta meirei" ni kansuru oboegaki [A note on "urgent imperatives"]. *Kyuushuu Daigaku Gengogaku Ronshyuu* [*Kyushu University Papers in Linguistics*] 35: 382–397.

Davis, Christopher. 2009. Decisions, dynamics and the Japanese particle *yo*. *Journal of Semantics* 26: 329–366.

Davis, Christopher. 2011. Constraining interpretations: Sentence final particles in Japanese. Doctoral dissertation, University of Massachusetts, Amherst.

von Fintel, Kai. 2004. Would you believe it? The King of France is back! Presuppositions and truth-value intuitions. In *Descriptions and Beyond*, ed. by Marga Reimer and Anne Bezuidenhout, 315–341. Oxford: Oxford University Press.

von Fintel, Kai, and Sabine Iatridou. 2017. A modest proposal for the meaning of imperatives. In *Modality across Syntactic Categories*, ed. by Ana Arregui, María Luisa Rivero, and Andrés Salanova, 288–319. Oxford: Oxford University Press.

Ihara, Shun. 2020. Decomposing directive strategies. Doctoral dissertation, Osaka University.

Ihara, Shun. 2021. Division of labor between semantics and pragmatics of canonical and non-canonical imperatives. *Gengo Kenkyu* 160: 155–182.

Ihara, Shun, and Yuya Noguchi. 2018. Imperatives with/without necessity. In *Japanese/Korean Linguistics 25*, ed. by Shin Fukuda, Mary Shin Kim, and Mee-Jeong Park, 103–116. Stanford: CSLI Publications.

Kaufmann, Magdalena. 2012. *Interpreting Imperatives*. New York: Springer.

Kaufmann, Magdalena, and Stefan Kaufmann. 2012. Epistemic particles and performativity. In *Proceedings of SALT 22*, ed. by Anca Chereches, 208–225. Washington, DC: Linguistic Society of America.

Kratzer, Angelika. 1991. Modality. In *Semantik: Ein Internationales Handbuch der Zeitgenössischen Forschung*, ed. by Arnim von Stechow and Dieter Wunderlich, 639–650. Berlin: de Gruyter.

Kratzer, Angelika. 2012. *Modals and Conditionals*. Oxford: Oxford University Press.

Noguchi, Yuya. 2016. Does ForceP determine all the illocutionary forces?: A case study of dictionary form imperatives in Japanese. In *Proceedings of FAJL 8: Formal Approaches*

to Japanese Linguistics, ed. by Ayaka Sugawara, Shintaro Hayashi, and Satoshi Ito, 75–85. Cambridge, MA: MITWPL.

Oshima, David Yoshikazu. 2011. On the functions of the Japanese discourse particle *yo* in declaratives. In *Formal Approaches to Semantics and Pragmatics: Japanese and beyond*, ed. by Elin McCready, Katsuhiko Yabushita, and Kei Yoshimoto, 251–271. New York: Springer.

Portner, Paul. 2004. The semantics of imperatives within a theory of clause types. In *Proceedings of SALT 14*, ed. by Robert B. Young, 235–252. Washington, DC: Linguistic Society of America.

Stalnaker, Robert. 1978. Assertion. *Syntax and Semantics* 9: 315–332.

Subatomic and plural homogeneity as exhaustification effects of different kinds*

Mathieu Paillé

University of Calgary, Calgary, Alberta, Canada
mathieu.paille@ucalgary.ca

Abstract

Homogeneity (all-or-nothing) effects are observed in both atoms and pluralities. In this paper, I compare two theories of homogeneity. The first is made for plural homogeneity and the second for subatomic homogeneity, but both capture the effect through existential lexical meaning paired with local exhaustification in positive sentences. I show that neither approach can be extended to the side of the paradigm it was not intended for. But relying on exhaustification for at least subatomic homogeneity is well motivated: all predicates belonging to taxonomies (rather than scales), whether displaying homogeneity effects or not, are interpreted as weak/strong in the same environments, and it is not clear what mechanism other than exhaustification could account for these facts in a united way. Hence, I maintain that homogeneity is an exhaustification effect, and suggest that plural and subatomic homogeneity are simply due to different kinds of local exhaustification.

1 Introduction

Discussion of homogeneity effects usually focuses on pluralities (1), but the effect holds within atoms as well (2) (e.g., Löbner 2000, Spector 2013, Križ 2015, 2019) with so-called 'summative' predicates—predicates that are true of an individual by virtue of being true of its parts.

(1) a. Adam saw the children.
≈ he saw **all** of the children
b. Adam didn't see the children.
≈ he saw **none** of the children

(2) a. The flag is green.
≈ **all** of the flag is green
b. The flag isn't green.
≈ **none** of the flag is green

In this paper, I compare two similar accounts of homogeneity, both of which posit existential lexical meaning that is strengthened in positive sentences but not negative ones. Bar-Lev (2021) proposes a theory of plural homogeneity (1) that obtains strengthening through the Inclusion of alternatives; Harnish (1976), Levinson (1983), and Paillé (2021) have a theory of summative predicates (2) that obtains strengthening through the Exclusion of alternatives. I take these theories in turn (sections 2–3) to show that neither cannot be extended to the side of the homogeneity paradigm it does not set out to explain. Then, in section 4, I defend that (2a) should be viewed as a strengthening effect: in particular, this makes it possible to collapse identical data around summative (3a) and non-summative (3b) predicates.

(3) a. The white flag is #(also) red.
b. This comedy is #(also) a tragedy.

*I would like to thank Bernhard Schwarz, Luis Alonso-Ovalle, Aron Hirsch, the audience at the 2022 Workshop on Homogeneity and Non-Maximality (online), and the reviewers for the 2022 Amsterdam Colloquium. This research was supported by a Vanier Canada Graduate Scholarship.

I therefore suggest that the Inclusion and Exclusion theories are both correct for a different part of the homogeneity paradigm (plural vs. subatomic homogeneity). What unites (1) and (2) is that both are exhaustification effects (and necessarily locally computed, as we will see), but the Exh(aust) operator (Chierchia et al. 2012) operates over different kinds of alternatives and, as a result, differs in whether it Includes or Excludes them.

2 An *Inclusion* account of *plural* homogeneity

Bar-Lev (2021) develops a theory of plural homogeneity based in the inclusion of parts; in plural predication, a predicate in a positive sentence is assigned the value 'true' for each part of a plurality, through a strengthening operation (Inclusion) carried out by Exh.

On this view, the meaning of plurals is existential: the plain meaning of *the children laughed* is that at least one laughed (assume there are two children, a and b).

(4) $[\![\text{The kids laughed}]\!] = 1$ iff $\mathsf{laughed}(a) \vee \mathsf{laughed}(b)$.

This existential meaning is due to an existential plural operator, ∃-PL:

(5) a. $[\![\exists\text{-PL}]\!] = \lambda D_{\langle et\rangle}.\lambda P_{\langle e,\, st\rangle}.\lambda x_e.\exists y \in D \cap Part_{\text{AT}}(x)[P(y) = 1]$. (Bar-Lev 2021:1062)
b. $Part_{\text{AT}}(x) = \{y : y \sqsubseteq_{\text{AT}} x\}$

The kids laughed has the LF in (6a), where the domain D is presented as a subscript on ∃-PL; (6a) obtains the meaning in (6b), which (given that $[\![\text{the kids}]\!] = a \oplus b$) is equivalent to (4).

(6) a. [The kids] [$\exists\text{-PL}_D$ laughed]. (Bar-Lev 2021:1062)
b. $[\![(6a)]\!] = 1$ iff $\exists y \in D \cap Part_{\text{AT}}([\![\text{the kids}]\!])[\mathsf{laughed}(y) = 1]$.

When (6a) is negated, this proposal immediately obtains the right truth conditions. For positive sentences, (6b) must be strengthened. Bar-Lev posits subdomain alternatives:

(7) ALT = {a laughed ∨ b laughed, a laughed, b laughed}

The sentence is exhaustified, but because the alternatives are not closed under conjunction and the alternatives 'a laughed' and 'b laughed' are not Innocently Excludable (Fox 2007), Exh excludes nothing. Instead, Bar-Lev uses the notion of Innocent Inclusion, according to which Exh asserts that non-excluded alternatives are *true*, insofar as this can be done consistently:

(8) **Innocent Inclusion procedure**: (Bar-Lev 2021:1067)
a. Take all maximal sets of alternatives that can be assigned true consistently with the prejacent and the falsity of all [innocently excluded] alternatives.
b. Only include (i.e., assign true to) those alternatives that are members in all such sets—the **Innocently Includable** alternatives.

Thus, when Exh takes (6a) as its prejacent, it asserts the alternatives 'a laughed' and 'b laughed'; this creates the meaning that all the children laughed.

At first glance, this account could be extended to subatomic homogeneity; there would be a non-plural operator akin to (5) referring to the subatomic parts of atoms. I assume that subatomic homogeneity is defined in terms of the arbitrary pieces (rather than the salient parts) of objects. For presentation, I will discuss a flag made up of two pieces, a and b, but this is an abstraction, as objects are made up of an infinity of (possibly overlapping) arbitrary pieces. On this view, the weak truth conditions in (9) would hold prior to exhaustification.

(9) $[\![\text{The flag is red}]\!] = 1$ iff $\mathsf{red}(a) \vee \mathsf{red}(b)$.

Including the subdomain alternatives 'a is red' and 'b is red' results in the meaning that the flag is all red; after all, a and b are standing in for an infinity of pieces covering the entire flag.

Despite these first appearances, trying to extend this Inclusion account to subatomic homogeneity leads to problems with data from conjoined colour predicates like (10).

(10) a. The flag is red **and** white.
b. The red flag is **also** white.

Positing the Inclusion of pieces to explain the quantificational strength of colour terms does not create the intuited meanings for (10). To see this, we must first appreciate that both sentences in (10) involve Boolean (intersective) conjunctive material managing to co-predicate two colour terms consistently (Paillé 2021).

(10b) is straightforwardly Boolean; additive particles have never been claimed not to be Boolean. But (10a) is less obvious. In fact, Krifka (1990) takes such examples to involve a non-Boolean conjunction; he assumes that colour terms are lexically universal, so the only way that (10a) might be consistent is if *and* predicates each adjective of a different part of the subject. A non-Boolean *and* makes it possible to predicate each conjunct of a different part of the subject (11a), resulting in consistent truth conditions (11c) even with lexically universal colour terms (11b).

(11) a. $[\![\text{and}]\!] = \lambda P.\lambda Q.\lambda x.\exists x', x''[x = x' \oplus x'' \wedge P(x') \wedge Q(x'')]$.
b. $[\![\text{red}]\!] = \lambda x.\forall y[y \sqsubseteq x \rightarrow \mathsf{red}(y)]$.
c. $[\![\text{(10a)}]\!] = 1$ iff $\exists x, x'[\text{the.flag} = x \oplus x' \wedge \mathsf{red}_\forall(x) \wedge \mathsf{white}_\forall(x')]$.

There is nothing wrong with the truth conditions in (11c); the question is whether they arise due to quantification by a non-Boolean *and*. Descriptively, with plural subjects, the availability of non-Boolean *and* means that the conjuncts can be made explicitly incompatible via *completely*:

(12) The flags are completely red and completely white.
≈ some of the flags are completely red, and the rest are completely white

If the truth conditions in (11c) arise due to quantification by *and*, it should also be possible to make the colour terms in (10a) explicitly incompatible—like (12), but with an atomic subject:

(13) #The flag is completely red and completely white.

This is inconsistent; Krifka's prediction did not hold up. Thus, it cannot be the case that there is a non-Boolean *and* (11a) available when the subject is atomic, as in (10a) and (13).

In sum, given the consistency of both examples in (10), a desideratum for theories of subatomic homogeneity is to capture that colour predicates are consistent (non-universal) when conjoined intersectively. Can the Inclusion theory of subatomic homogeneity meet this challenge? For this theory to stand a chance, let's start by assuming that colour terms are lexically existential (14), so that they can be consistently conjoined through a Boolean *and* or *also*.

(14) $[\![\text{red}]\!] = \lambda x.\exists y[y \sqsubseteq x \wedge \mathsf{red}(y)]$.

Prior to exhaustification, on the Inclusion theory, (10a) has the truth conditions in (15a) and the alternatives in (15b).

(15) a. $[\![$The flag is red and white$]\!]$
$= 1$ iff $(a$ is red$_\exists \wedge a$ is white$_\exists) \vee (b$ is red$_\exists \wedge b$ is white$_\exists)$

b. $\text{ALT} = \left\{ \begin{array}{c} a \text{ is red}_\exists \text{ and white}_\exists \vee b \text{ is red}_\exists \text{ and white}_\exists, \\ a \text{ is red}_\exists \text{ and white}_\exists, \\ b \text{ is red}_\exists \text{ and white}_\exists \end{array} \right\}$

When (15a) is exhaustified, none of the alternatives are excludable. But attempting to include the alternatives does not create the desired meaning. Recall that a and b stand in, for simplicity of presentation, for two arbitrary subatomic pieces of the flag; but in reality, there are an infinite amount of such pieces, which together cover the entire flag; and these can overlap. By including the alternatives in (15b), the meaning obtained is that for any arbitrary piece of the flag, that piece has both a partly red piece and a partly white piece. These pieces too must have a partly red piece and a partly white piece, and so on—all the way down. Every arbitrary piece would have to be divisible into differently-coloured pieces, which themselves are divisible in the same way. This is not the intuited meaning at all; in fact, such an object is impossible.

Perhaps this is such a problematic meaning that the alternatives in (15b) are simply not includable. If so, (15) is still problematic. The meaning of (10a) would simply be that there is a piece of the flag which has a red piece and also has a white piece. The colour terms remain existential. This is far from the intuition for (10a), which is that the flag is *only* red and white; the colour terms are stronger than merely existential.

For this reason, the Inclusion theory, while capable of dealing with plural homogeneity, should not be carried over to subatomic homogeneity.

3 An *Exclusion* account of *subatomic* homogeneity

A rather different account of homogeneity, also based in exhaustification, is proposed by Harnish (1976), Levinson (1983), and Paillé (2021) for subatomic homogeneity. Colour predicates are taken to be existential here as well (16a), obtaining the meaning of negative sentences immediately. Positive sentences are strengthened not through the Inclusion of all parts, but the Exclusion of other existential colour terms (16b).

(16) a. $[\![\text{red}]\!] = \lambda x.\exists y[y \sqsubseteq x \wedge \mathsf{red}(y)]$.
b. $[\![\text{Exh}_{\text{ALT}}$ [the flag is red]$]\!]$
$= 1$ iff the flag is red$_\exists \wedge$ the flag is not white$_\exists \wedge$ the flag is not green$_\exists \wedge \ldots$

(16b) does not make *red* semantically universal. Semantically, it only negates that the flag has other colours; but given our conceptualization of surfaces, this is pragmatically strengthened to mean that the flag is entirely red. Thus, *red* is 'semantically exclusive, pragmatically universal.'

This Exclusion account deals with conjunctions effortlessly. An Exh operator scoping above the entire conjunction will only exclude non-asserted colour terms; Exh does not exclude alternatives entailed by its prejacent (Chierchia et al. 2012).

(17) $[\![\text{Exh}_{\text{ALT}}$ [the flag is red and white]$]\!]$
$= 1$ iff the flag is red$_\exists \wedge$ the flag is white$_\exists \wedge$ the flag is not green$_\exists \wedge \ldots$

The account can also deal with the additive data (10b), since additives have independently been argued to weaken exhaustification (Bade 2016; see Paillé to appear on colour terms specifically).[1]

[1]The Exclusion account can also capture the truth-value gap intuited in non-homogeneous situations (see e.g. Löbner 2000 and Križ 2015) by adopting the trivalent Exh proposed by Bassi et al. (2021).

But the Exclusion theory cannot be extended to plural homogeneity. It crucially relies on world knowledge, so that 'partly red, and no other colour' is strengthened to meaning 'entirely red.' Nothing of the sort holds for pluralities; world knowledge does not dictate anything about the parts of pluralities. Translating the above Exclusion account to pluralities would involve claiming that plural predication is lexically existential, but is strengthened via the exclusion of conceptually related predicates:

(18) $[\![\text{Exh}_{\text{ALT}}\ [\text{the children are singing}_{\exists}]]\!] = 1$ iff $\begin{cases} \text{there is a child who is singing} \wedge \\ \text{there is no child who is dancing} \wedge \\ \text{there is no child who is} \ldots \end{cases}$

This both creates entailments that are straightforwardly not intuited (at least without contrastive focus on *singing*), and does not actually make the basic existential meaning 'pragmatically universal': it is possible for members of a plurality to be doing nothing. That is, the truth conditions in (18) are compatible with only one child singing, and the others doing nothing.

A reviewer for the Amsterdam Colloquium suggests that the Exclusion account could work by constraining the alternatives in (18) to mutually exclusive predicates; *sing* would not have *dance* as an alternative, but it would have *talk* and *keep quiet*, for example. If at least one child in the plurality is singing, none is talking, and none is keeping quiet, they must all be singing. However, on this proposal, we lose the ability to explain the *subatomic* homogeneity data, where alternatives are crucially *consistent* predicates (the colour terms are lexically existential). If only *inconsistent* predicates were alternatives, colour terms would not meet this requirement, and would therefore not be alternatives to one another.

4 An argument for homogeneity as exhaustification

An important benefit of the Exclusion account of subatomic homogeneity is that it can also capture the strength of non-summative predicates. Indeed, the Exclusion account of colour terms is part of a broader theory of predication (Paillé 2020); I claim it is a *general* property of predicates to exclude conceptually related predicates (that is: taxonomic sisters). This is motivated by data like (19). On the view that additive particles act to weaken or remove unwanted exhaustification (Krifka 1998; Sæbø 2004; Bade 2016), (19) shows that the meaning of many different predicates involves exhaustification. Unlike colour terms, the predicates in (19) are not summative; a fork is not a fork by virtue of its parts being forks, for example.

(19) a. A tragicomedy is a tragedy that is #(also) a comedy. (Paillé 2020)
b. A spork is a fork that is #(also) a spoon.
c. Some federal responsibilities are #(also) provincial.
d. Futons are couches that are #(also) beds.
e. Are any derivational morphemes #(also) inflectional?
f. Some left-wing ideas are #(also) right-wing.

Given (19), it must be that predicates are always exhaustified vis-à-vis their taxonomic sisters:

(20) $[\![\text{This}\ [\text{Exh}_{\text{ALT}}\ \text{tragedy}]\ \text{is a}\ [\text{Exh}_{\text{ALT}}\ \text{comedy}]]\!]$
$= 1$ iff this $\begin{pmatrix} \text{tragedy} \ \& \\ \text{not a comedy} \ \& \\ \text{not an epic} \end{pmatrix}$ is a $\begin{pmatrix} \text{comedy} \ \& \\ \text{not a tragedy} \ \& \\ \text{not an epic} \end{pmatrix}$ $\Rightarrow$ contradiction

As such, the Exclusion account of subatomic homogeneity makes it possible to view this

paradigm as part of a broader phenomenon of strengthening in predication; this is motivated empirically by the identical behaviour of colour terms with additive particles (cf. (10b)):

(21) a. The white flag is #(also) green.
b. $[\![\text{The } [\text{Exh}_{\text{ALT}} \text{ white}] \text{ flag is } [\text{Exh}_{\text{ALT}} \text{ green}]]\!]$
$= 1$ iff the $\begin{pmatrix} \text{white}_\exists \ \& \\ \text{not green}_\exists \ \& \\ \text{not red}_\exists \end{pmatrix}$ is $\begin{pmatrix} \text{green}_\exists \ \& \\ \text{not white}_\exists \ \& \\ \text{not red}_\exists \end{pmatrix}$ $\Rightarrow$ contradiction

In sum, the Exclusion account of homogeneity can collapse data concerning the quantificational strength of summative predicates with the 'conceptual' strength of the predicates in (19). This advantage of the Exclusion theory makes it unappealing to throw it out—especially in favour of a theory of homogeneity that refers explicitly to part-structure, and is therefore impossible to carry over to the predicates in (19). Since we must keep the exhaustivity account of subatomic homogeneity as the only currently available option that can also explain (19), it seems best to claim that plural homogeneity is an exhaustification effect, too. On this view, plural and subatomic homogeneity at least have substantial common ground.

5 Conclusion: homogeneity as a local exhaustivity effect

We have considered two similar accounts of homogeneity that use Exh and weak lexical meaning to derive the paradigm: an Inclusion account made for pluralities and an Exclusion account made for atoms. Neither can be extended to the other half of the homogeneity paradigm. However, it remains that the two approaches share an important component: they involve weak lexical meaning together with covert exhaustification in positive sentences. This may be all that formally unites plural and subatomic homogeneity.

As exhaustification effects, plural and subatomic homogeneity must be claimed to only be computed locally. For colour terms, this claim is required for (21) as well as other data like (22a). Indeed, for (22a), a global Exh would not make *red* pragmatically universal; the truth conditions in (22b) are compatible with every flag being only partly red, as long as there is no other colour such that all flags are partly of that colour.

(22) a. Every flag is red.
$\approx$ every flag is entirely red
b. $[\![\text{Exh}_{\text{ALT}} \text{ [every flag is red]}]\!] = 1$ iff $\begin{cases} \text{every flag is red}_\exists \wedge \\ \neg[\text{every flag is white}_\exists] \wedge \\ \neg[\text{every flag is green}_\exists] \end{cases}$

It must therefore be claimed for (22a) that a local Exh, below *every*, is the only parse available. A similar locality requirement must be posited for the Inclusion account of plural homogeneity: positive clauses are intuited as involving universal quantification even in downward-entailing environments (Križ 2015; Križ & Spector 2021) (23), so it must be claimed that Exh is necessarily located below the DE operator.

(23) If you solve the problems, you will pass the exam. (Križ 2015:27)
$\approx$ 'If you solve **all** the problems, you will pass the exam.'

As such, subatomic and plural homogeneity are united specifically in being necessarily local exhaustification effects. How to derive such a locality constraint on Exh is left for other work.

References

Bade, Nadine. 2016. Obligatory presupposition triggers in discourse: Empirical foundations of the theories *Maximize Presupposition* and *Obligatory Implicatures*. Doctoral dissertation, University of Tübingen, Tübingen.

Bar-Lev, Moshe E. 2021. An implicature account of Homogeneity and Non-maximality. *Linguistics and Philosophy* 44:1–53.

Bassi, Itai, Guillermo Del Pinal, & Uli Sauerland. 2021. Presuppositional exhaustification. *Semantics and Pragmatics* 14. Early access version.

Chierchia, Gennaro, Danny Fox, & Benjamin Spector. 2012. Scalar implicatures as a grammatical phenomenon. In *Semantics: An international handbook of natural language meaning*, ed. Claudia Maienborn, Klaus von Heusinger, & Paul Portner, volume 3, 2297–2331. Berlin: De Gruyter Mouton.

Fox, Danny. 2007. Free choice and the theory of scalar implicatures. In *Presupposition and implicature in compositional semantics*, ed. Uli Sauerland & Penka Stateva, 71–120. Basingstoke, Hampshire (UK): Palgrave Macmillan.

Harnish, Robert. 1976. Logical form and implicature. In *An integrated theory of linguistic ability*, ed. Thomas G. Bever, Jerrold J. Katz, & D. Terence Langendoen, 313–391. New York: Thomas Cromwell.

Krifka, Manfred. 1990. Boolean and non-Boolean 'and'. In *Papers from the Second Symposium on Logic and Language*, ed. László Kálmán & László Pólos, 161–188. Budapest: Akadémiai Kiadó.

Krifka, Manfred. 1998. Additive particles under stress. In *Proceedings of SALT 8*, ed. Devon Strolovitch & Aaron Lawson, 111–129. Ithaca, NY: Cornell University.

Križ, Manuel. 2015. Aspects of homogeneity in the semantics of natural language. Doctoral dissertation, University of Vienna, Vienna.

Križ, Manuel. 2019. Homogeneity effects in natural language semantics. *Language and Linguistics Compass* 13:1–17.

Križ, Manuel, & Benjamin Spector. 2021. Interpreting plural prediction: homogeneity and non-maximality. *Linguistics and Philosophy* 44:1131–1178.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

Löbner, Sebastian. 2000. Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* 23:213–308.

Paillé, Mathieu. 2020. The distribution of controlled exhaustivity. In *Proceedings of SALT 30*, ed. Joseph Rhyne, Kaelyn Lamp, Nicole Dreier, & Chloe Kwon, 843–860.

Paillé, Mathieu. 2021. Exhaustivity and the meaning of colour terms. In *Proceedings of WCCFL 38*, ed. Rachel Soo, Una Y. Chow, & Sander Nederveen, 334–344. Somerville, MA: Cascadilla Press.

Paillé, Mathieu. to appear. On additives' interaction with exhaustivity: the view from negative continuations. In *Proceedings of NELS 52*.

Sæbø, Kjell Johan. 2004. Conversational contrast and conventional parallel: Topic implicatures and additive presuppositions. *Journal of Semantics* 21:199–217.

Spector, Benjamin. 2013. Homogeneity and plurals: From the Strongest Meaning Hypothesis to supervaluations. Sinn und Bedeutung 18: University of the Basque Country, Vitoria-Gasteiz. Available at: https://ehutb.ehu.es/uploads/material/Video/3289/Sinn18_01.pdf.

Numerous relative clauses
permutation invariance, anti-restrictiveness, triviality*

Lorenzo Pinton

Massachusetts Institute of Technology
lpinton@mit.edu

Abstract

In this paper we study the different patterns that *gather*-like and *numerous*-like predicates present in restrictive relative clauses. We argue that *numerous* is generally bad in restrictive relative clauses, because when it applies to a predicate that denotes a complete join semilattice, triviality arises with the definite plural. As a consequence, we predict that restrictive *numerous* is fine if it applies to predicates that do not give rise to complete join semilattices. This is indeed the case for complex predicates like *gathered students*, whenever the collective predicate *gather* is not pluralized.

1 Introduction

The present paper aims at providing an analysis to some puzzles linked to restrictive uses of *gather*(-like) and *numerous*(-like) predicates. The first puzzle arises from the contrast shown by the following sentences:[1]

(1) a. Jack only talked to the students that gathered.
b. #Jack only talked to the students that were numerous.

Whereas (1a) is a felicitous sentence in which the predicate *students* has been modified by *gathered* by means of a relative clause, (1b) - the result of substituting the predicate *gather* with *be numerous* - is somehow deviant.[2] A preliminary question we have to ask is whether the general infelicity of restrictive relative clauses involving the predicate *numerous* is specific to relative clauses or is actually linked to restrictiveness *per se*. With respect to this issue, the following data from Italian is particularly helpful:

(2) a. Ho parlato con i numerosi studenti.
have.1SG talked with the.PL numerous.PL student.PL
I have talked with the numerous students.
b. ??Ho parlato con gli studenti numerosi.
have.1SG talked with the.PL student.PL numerous.PL
I have talked with the numerous students.

*The main examples and the problems they constitute, together with important suggestions toward a solution, were proposed by Martin Hackl in a problem set for the course 'Introduction to Semantics'. Thanks also to all the participants to 'Introduction to Semantics', 'Workshop in Linguistics' and the 33rd European Summer School in Logic, Language and Information for crucial feedback on an early version of this work, in particular Martin Hackl, Keely New, Jad Wehbe, Omri Doron, Dean McHugh, Patrick Elliott, and six anonymous reviewers.

[1] All our examples include *only* in the attempt to make sure that the relative clause is interpreted as a restrictive relative clause.

[2] There are some speakers that upon reflection manage to find (1b) acceptable 'imagining a scenario in which there are different groups of students'. We claim that in this case *students* is contextually interpreted as *gathered students*. We unfortunately don't have space to discuss whether this interpretation can be due to a cover analysis or to an implicit *gathered*. But we believe that for the intended purposes they are equivalent, even though the cover analysis might result in additional assumptions on the application of relative clauses.

The minimal pair above exploits a typical Romance phenomenon: modifiers in pre-nominal positions are linked to non-restrictive interpretations, while post-nominal modifiers are ambiguous between the two readings [Mor08, Cin03].[3] The contrast between (2a) and (2b) signals then that the infelicity of the predicate *essere numeroso / to be numerous* is really tied to restrictive interpretations in general, rather than limited to (restrictive) relative clauses. There must be something that blocks *numerous* from being restrictive, namely from selecting a specific subset of a given set.[4]

The contrast observed in (1) becomes even more compelling when compared with the second puzzle we will discuss, which arises from the following sentence:

(3) Jack only talked to the gathered students that were numerous.

This crucial data that seems to contradict the generalization we stated above: (3) shows that *numerous*-restrictive relative clauses can be felicitous; in particular, they are perfectly fine if the predicate they combine with is a complex predicate, in our case if *student* has already been modified by some other predicate, *gather*. We will show that a specific internal structure is required for these predicates, and we will derive the felicity of (3) as a prediction of our analysis for the infelicity of (1b).

To summarize the goal of this paper, we can take away two main observations and two questions, respectively, from the examples presented insofar:

1. the modifier *numerous* is generally fine as non-restrictive but not as restrictive; what is preventing it from being restrictive?

2. there exist cases in which *numerous* is fine in restrictive position; what is the difference with the other cases?

To answer the first question, we claim that restrictive readings are blocked whenever a *permutation invariant* predicate like *numerous* applies to complete join semilattices. This property leads to redundancy: the individual resulting from the application of the relative clause has no potential to differ (be restricted) from the individual that would have been denoted without the relative clause.[5]

To answer the second question, we will build on the intuition that (3) involves several groups of gathered students. We will argue that while most complex predicates (like *smiling students*) preserve a complete join semilattice structure, others (like *gathered students*) have the possibility to give rise to other structures. In particular, we will claim that *gathered* changes the complete join semilattice structure of *students* when it is not pluralized. We will assume that pluralization comes from a syntactic operator, the star operator [Lin87], in line with [Ste98] and [BS00]. On the one hand, we assume that pluralization can apply to every base-generated and complex predicate; on the other hand, we assume pluralization to be optional, or, more precisely, omitted if its application leads to triviality. Let us now start from an introduction of the relevant properties of *gather*-like and *numerous*-like predicates that we will borrow from the existing literature to provide an analysis to our puzzles.

[3] Crucially, we do not commit here to any specific syntactic or semantic analysis of the contrast between the two interpretations, beyond the idea that restrictive interpretations select a subset of the predicate they combine with.

[4] We don't delve here into the thorny question of the proper semantics for restrictiveness and non-restrictiveness, and we just treat restrictiveness as intersection of predicates.

[5] Our solution is marginally linked to the *anti-monotonicity constraint* proposed in [Ami21]. However, we do believe that the interplay between the structure of plurals and *permutation invariance* is the real notion at play here, and monotonicity is just a related consequence that shows up most but not all of the times.

2 *Gather*, *be numerous*, and their relevant properties

Both *gather* and *be numerous* can be classified as collective predicates, for they only accept plural individuals in their domain. However, it has been known for a while now that a more fine-grained distinction is needed within the class of collective predicates, since *gather* and *numerous* give rise to different patterns when combined with plural quantifiers [Kro74, D$^+$87, Cha10]. These patterns resemble the contrast we presented in (1), since only so-called *gather*-type predicates can felicitously combine with plural quantifiers, as displayed in (4) and (5).

(4) a. All the students gathered.
b. #All the students are numerous.

(5) a. Most of the students gathered.
b. #Most of the students are numerous.

A couple of intuitions on the properties that are linked to this problem are generally shared by the literature, and, following [Kuh20], they can be summarized as follows:

1. *numerous*-like predicates denote properties of groups as such;
2. *gather*-like predicates license 'distributive sub-entailments'.

Property 1 states that *numerous*-like predicates do not hold (or are not guaranteed to hold) for subparts of what they are predicated of. However, the pivotal notion we want to stress here, is that these predicates are 'blind to differences', they denote properties of groups but they cannot distinguish whether a certain atom instead of another is in the plurality they are predicated of. We call this property *permutation invariance*. Take for example the case of *numerous*: it only checks whether a given plurality has a certain cardinality, regardless of the properties of the specific members that constitute such plurality. To say that $a \oplus b$ are a *numerous* plurality, we are not concerned with any of the properties of a and b, it suffices to say that the plurality $a \oplus b$ exists, i.e. we are only interested in the fact that we can group a and b together. For instance, if we substituted a with c, the numerosity of the plural individual would not change. In contrast, when *gather* is predicated of $a \oplus b$, it is crucial that a and b participated in a gathering event, and substituting b for c, will not necessarily guarantee that $a \oplus c$ gathered too. In section §3 will discuss how this property play a crucial role in blocking restrictive interpretations.

With respect to property 2, [Kuh20] traces back the distinction between the two classes of collective predicates to a mass/count distinction. In particular, in the same way every part of something that is water is water itself, *gather*-like predicates have the crucial property of holding for any plural subpart of the individual they are predicated of.[6] However, here we are interested in the reverse property, quickly mentioned by [Kuh20]. As a collective predicate, *gather* denotes pluralities by definition, but when it is not pluralized it does not hold for the sums of such pluralities. This property plays a crucial role in creating a set of plural individuals that is not a complete join semilattice, and therefore does not lead restrictive modification by *numerous* to triviality. Let us now introduce the details on how these properties explain the behaviour of *numerous*-relative clauses.

[6] [Distributive sub-entailments]If a predicate holds for a plurality x, then it holds for any y s.t. $y \leq x$ and $\neg Atom(y)$. 'ε-bounded divisiveness' is a variant of this principle also used in the literature [Kuh20]. The general idea of 'distributive sub-entailments' was first proposed in [D$^+$87] and Kuhn has recently improved the analysis to account for several over and undergenerations.

3 The analysis, starring: most verbs but not *gather*

3.1 Permutation invariance, i.e. anti-restrictiveness, i.e. triviality

Recall our initial example (1b), repeated here as (6):

(6) # Jack only talked to the students that were numerous.

In this section, we aim at showing how, because of the algebraic properties of plurals, whenever *numerous* tries to restrict a pluralized predicate, it inevitably ends up receiving a non-restrictive interpretation, which results in ungrammaticality when this interpretation cannot be derived by any parse of a given sentence. Consider the denotation for *students* as the complete join semilattice structure used to analyse plurals due to [Lin87] (obtained interpreting the plural morpheme *-s* as the star operator[7]), together with the following definition for *numerous*:

(7) $[\![numerous]\!]^{g_c} = \lambda x : x \in dom(|\ |).|x| \geq$ the contextually supplied numerosity standard

In simple terms, *numerous* are all those plural individuals with a cardinality above a certain threshold. A visual representation is showed in Fig.1, where the numerosity threshold has been set to 2, for the sake of simplicity.

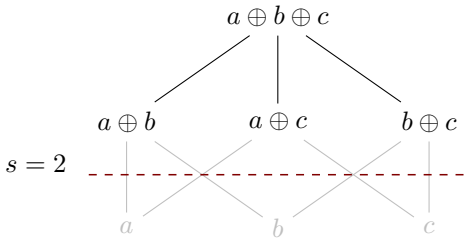


Figure 1: $[\![numerous\ {}^*students]\!]^{g_c} = \{a \oplus b, a \oplus c, b \oplus c, a \oplus b \oplus c\}$

With no additional contextual restriction on how we can pick plural individuals, *any* group with cardinality 2 or more is going to satisfy *be numerous*.[8] This means that for any x such that x is an atom, there is at least one y such that y is a numerous individual and $y \succeq x$. We call this property *anti-restrictiveness*, and it is a property that predicates like *be numerous* / *be few in number* / *be exactly 3 in number* / etc. carry whenever they are applied to a complete join semilattice. The composite predicate resulting from the application of the *numerous*-relative clause to the original predicate of the NP (*students* in our example) has then to be pluralized (i.e. closed under sum formation via star), to ensure composition with the definite article, to satisfy the existence of a unique maximal element.[9] However, if any atom is part of at least one *numerous* individual, the maximal element resulting from the application of star to *students*

[7] Where ${}^*h$ is the smallest function s.t. $\forall x[h(x) = 1 \rightarrow^* h(x) = 1]$ and $\forall x, y[{}^*h(x) =^* h(y) = 1 \rightarrow^* h(x \oplus y) = 1]$ (Martin Hackl, lecture notes).

[8] If we wanted to claim that $a \oplus b$ from the previous figure is a numerous individual we wouldn't be able to hold that $a \oplus b$ is a numerous individual without admitting that $a \oplus c$, $b \oplus c$, and $a \oplus b \oplus c$ are numerous individuals too.

[9] Assuming for the definite article the following meaning adapted from [Sha80]: $[\![the]\!]^{g_c} = \lambda F : F_{\langle e,t\rangle} \& \exists x[F(x) \& \forall y[F(y) \rightarrow y \preceq x]. \iota z[F(z) \& \forall y[F(y) \rightarrow y \preceq z]$. Pluralization after the relative clause composition is vacuous for *numerous*, since it already includes the maximal element (because of monotonicity with respect to parthood), but it would be required for other predicates of this sort like *be few in number* and *be exactly 3 in number*.

that are numerous will just be the maximal element of *students*, i.e. the sum of every atomic individual in the predicate set. The supposed restriction via *numerous* results therefore in *triviality*.

3.2 Collective predicates and structural modifications

As stated above, *numerous*-like predicates will end up in triviality whenever they are applied to complete join semilattices. This means that they would result in ungrammaticality whenever they are applied to pluralized predicates (*students)*, or to intersections of pluralized predicates (*smiling students*). Note that, if we try to not pluralize these distributive predicates, *numerous* would still fail to apply, this time because, it being a plural predicate, it cannot be applied to atomic entities.[10] On the other hand, we do expect *numerous*-like predicates to be compatible with predicates that can provide them with suitable individuals (plural entities), without generating a complete join semilattice. This is the case for collective predicates when they are not pluralized: they apply to plural entities by definition but not necessarily to their sums. In the same way that a non-pluralized distributive predicate like *smile* would have in its denotation only the atomic individuals that smile, a predicate like *gather* would only have the groups that gathered together (and their sub-entailments). For instance, if the students gathered in one room and the professors in another, the sum of students and professors would not be in the denotation of unpluralized *gather*. In other words, modifications with unpluralized *gather*-like predicates is not semilattice preserving. As a consequence, the maximality problem with restrictive *numerous* does not arise. This is why (3), repeated here as (8), is felicitous:

(8) Jack only talked to the gathered students that were numerous.

The scenario described by speakers when confronted with (8) is one in which there are multiple student groups, some of them being numerous, and some of them being not. Consider then an instance of this scenario in which a, b, c, e and f are students that gathered in two groups: $a \oplus b \oplus c$ and $e \oplus f$ and take '3' as the contextually supplied cardinality necessary for *numerous* to hold. If we then apply *numerous* to the meaning of *gathered students* (in Fig.2 on the next page), where *gather* is unpluralized, the result will give us $a \oplus b \oplus c$ as the only numerous individual in the set.

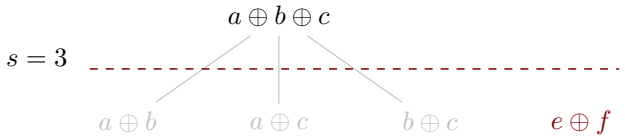


Figure 2: $[\![\textit{gathered *students}]\!]^{g_c} = \{a \oplus b, a \oplus c, b \oplus c, e \oplus f, a \oplus b \oplus c\}$
$[\![\textit{gathered *students that are numerous}]\!]^{g_c} = \{a \oplus b \oplus c\}$

In this structure restriction manages to go through, selecting a numerous individual that excludes $e \oplus f$: there is no numerous y such that $e \preceq y$ or $f \preceq y$. As a consequence, when we pluralize *gathered students that are numerous* to make sure we meet the definite article presupposition, *the* will pick up $a \oplus b \oplus c$, and not $a \oplus b \oplus c \oplus e \oplus f$. This time the relative clause successfully restricts the predicate: the maximal individual post-relative clause is strictly smaller

[10] And this might be the reason why speakers object that 'singular students cannot be numerous' when confronted with (1b). Probably pluralization is removed to avoid triviality, but a mismatch between categories of entities (singular vs plural) arises.

than the maximal individual pre-relative clause. The assumption that permits this machinery is that pluralization can apply to any node in which a new predicate is formed (e.g. *gather*, *gathered students*, *gathered students that are numerous*, ...), and can be freely omitted to avoid triviality. The LF of (8) would therefore be the following, where pluralization takes place on *students*, to compose with the collective predicate *gather*, and before the definite article, to meet its presupposition in case there are multiple groups of gathered students that were numerous.

(9) Jack only talked to [the [* [[[gathered] [* [students]]] [that were numerous]]]].

4 Conclusions and Future Work

In this work we provided an explanation to the contrast observed in (10a) and (10b):

(10) a. #Jack only talked to the students that were numerous.
b. Jack only talked to the gathered students that were numerous.

We claimed that the infelicity of (10a) is due to the permutation invariance of *numerous*-like predicates when applied to certain structures. Any atom of the predicate to which *numerous* applies cannot be excluded if such atom appears in every layer of the complete join semilattice generated by the predicate's pluralization. This implies that any proper restriction is forbidden, given that the maximal element necessary for the definite article *the* will be the sum of every single atom of the original predicate. We then explained how (10b) is a correct prediction of our theory, provided that unpluralized collective predicates can denote pluralities without being closed under sum formation. The underlying assumption of our system is that pluralization can apply to any node in which a new predicate is formed. The application of star in felicitous sentences, like (10b), is driven by the simultaneous satisfaction of two semantic and logical properties: meeting the presupposition of the definite article, while avoiding triviality. Such reading is impossible for infelicitous sentences like (10a), since removing pluralization to avoid triviality leads to a mismatch between the atomic individuals of distributive predicates and the plural individuals to which *numerous*-like predicates can apply. This solution has certain repercussions on the relation between the star operator and the plural morpheme that we would like to explore more in the future. For example, felicitous restrictive applications of *numerous*-relative clauses are found (as expected) with collective nouns. Consider the following example:

(11) We organized a competition only for [the [* [groups [that were numerous]]]].

It is clear that either we abandon the identification of the star operator with the plural morpheme -s, or we assume that the plural on *groups* (and possibily on *were*) morphologically attaches at a later stage than the one on *students* in (9). Along with this question, for future research we would like to fully explore the advantages of our approach compared to a possible alternative analysis based on covers, in the style of [Sch96]. While covers might be an appealing way to model the reading in which *gathered students* denotes multiple groups of students, it seems that such an approach would still require assumptions on pluralization similar to those made here. In particular, the cover approach could face some challenges in explaining how relative clauses compose with denotations resulting from non-pluralized covers, while still passing on pluralized meanings to the definite article. Finally, in the near future we would like to extend our solution to the other partitive constructions that present a contrast between *numerous*-like and *gather*-like predicates: plural quantifiers and non-proportional partitive constructions.

References

[Ami21] Omri Amiraz. A constraint on co-occurrence of partitive quantifiers and gradable predicates. In *Proceedings of Sinn und Bedeutung*, volume 25, pages 55–72, 2021.

[BS00] Sigrid Beck and Uli Sauerland. Cumulation is needed: A reply to winter (2000). *Natural language semantics*, 8(4):349–371, 2000.

[Cha10] Lucas Champollion. Parts of a whole: Distributivity as a bridge between aspect and measurement. philadelphia, pa: University of pennsylvania dissertation, 2010.

[Cin03] Guglielmo Cinque. The dual source of adjectives and xp-vs. n-raising in the romance dp. *Incontro annuale di dialettologia, Padua*, 26, 2003.

[D$^+$87] David Dowty et al. Collective predicates, distributive predicates, and all. In *Proceedings of the 3rd ESCOL*, pages 97–115. (Eastern States Conference on Linguistics), Ohio State University Ohio, 1987.

[Kro74] Anthony S Kroch. *The semantics of scope in English.* PhD thesis, Massachusetts Institute of Technology, 1974.

[Kuh20] Jeremy Kuhn. Gather/numerous as a mass/count opposition. *Natural Language Semantics*, 28(3):225–253, 2020.

[Lin87] Godehard Link. Generalized quantifiers and plurals. In *Generalized quantifiers*, pages 151–180. Springer, 1987.

[Mor08] Marcin Morzycki. Nonrestrictive modifiers in non-parenthetical positions. *McNally & Kennedy, Adverbs and Adjectives*, 2008.

[Sch96] Roger Schwarzschild. *Pluralities*, volume 61. Springer Science & Business Media, 1996.

[Sha80] Richard Sharvy. A more general theory of definite descriptions. *The philosophical review*, 89(4):607–624, 1980.

[Ste98] Wolfgang Sternefeld. Reciprocity and cumulative predication. *Natural language semantics*, 6(3):303–337, 1998.

Diversity with Universality*

Sonia Ramotowska[1], Paul Marty[2], Jacopo Romoli[1], Yasutada Sudo[3], and Richard Breheny[3]

[1] Heinrich-Heine-Universität Düsseldorf
[2] L-Università ta' Malta
[3] University College London

Abstract

This paper investigates the 'Diversity' inferences (D-INFERENCES) arising from disjunction embedded in the scope of a universal quantifier, e.g., *Every X is A or B* suggests *Some Xs are A* and *Some Xs are B*. It has previously been claimed (i) that D-INFERENCES are independent from 'Negative Universal' inferences (NU-INFERENCES), which are the negations of *Every X is A* and *Every X is B*, but (ii) that for disjunction in the scope of a universal modal the D-INFERENCES cannot be observed independently of the NU-INFERENCES ([3, 7, 9, 13]). Experiment 1 tested the availability of D-INFERENCES in the absence of NU-INFERENCES for the determiner *every* and the epistemic modal *must*. Experiment 2 followed up on Experiment 1 by testing the same two quantifiers, only this time the modal *must* expressed deontic necessity. The results show that, for both types of quantifiers, D-INFERENCES could be derived independently of NU-INFERENCES. While the results for *every* essentially replicate those reported in [7], the results for *must* are new and go against the aforementioned claim (ii). In addition, the response time results from both experiments show that D-INFERENCES are associated with response delay effects in the opposite direction to those observed for regular scalar implicatures in similar tasks ([4, 5]). We argue that these findings about the time course of D-INFERENCES raise a new challenge for an implicature-based approach to these inferences.

1 Introduction

A sentence like (1), where disjunction occurs in the scope of a nominal universal quantifier, gives rise to the inference in (1-a), schematised in (1-b). Inferences like (1-a) are generally referred to as 'Diversity' or 'Distributive' inferences. We will use the more compact label 'D-INFERENCES'.

(1) Every visible box contains a blue ball or a yellow ball. $\forall x(\mathsf{A}x \vee \mathsf{B}x)$

a. $\rightsquigarrow$ *Some visible box contains a blue ball and some contains a yellow ball*

b. $\exists x\mathsf{A}x \wedge \exists x\mathsf{B}x$

While the existence of D-INFERENCES is uncontroversial, the status and source of these inferences is still very much debated. Traditionally, they are considered scalar implicatures, derived through negating the alternatives to (1) in (2-a) and (2-b) ([8, 12]).

(2) a. Every visible box contains a blue ball. $\forall x(\mathsf{A}x)$

b. Every visible box contains a yellow ball. $\forall x(\mathsf{B}x)$

*We would like to thank Maria Aloni, Moysh Bar-Lev, Itai Bassi, Kyle Blumberg, Ivano Ciardelli, Marco Degano, Melissa Fusco, Paloma Jeretic, Matt Mandelkern, Paolo Santorio, Florian Schwarz, and Benjamin Spector. This research was supported by the Leverhulme Trust Grant RPG-2018-425

This account of D-INFERENCES makes a straightforward prediction: D-INFERENCES should always arise in combination with the additional 'Negative Universal' inferences (henceforth, NU-INFERENCES) in (3-a), schematised in (3-b).

(3) a. $\rightsquigarrow$ *Not all visible box contain a blue ball and not all of them contain a yellow ball*
b. $\neg\forall x \mathsf{A}x \wedge \neg\forall x \mathsf{B}x$

There is experimental evidence that challenges this prediction ([7]): Sentences like (1) were robustly judged as true, when their D-INFERENCES were true while their NU-INFERENCES were false (e.g., in situations where all of the visible boxes contain a yellow ball and some of them contain also a blue one), and rejected when their D-INFERENCES and NU-INFERENCES were both false (e.g., in situations where all of the visible boxes contain a yellow ball and none of them also contains a blue one). These results suggest that D-INFERENCES can arise independently of NU-INFERENCES. In fact, no evidence for the presence of NU-INFERENCES was found.

In addition to this finding, it has also been considered, based on introspective judgments, that for disjunction in the scope of a modal quantifier the D-INFERENCE cannot be observed independently of the NU-INFERENCE ([3, 7, 9, 13]). For example, the D-INFERENCE associated with a sentence like (4) is considered to always arise with its NU-INFERENCE, (4-b).

(4) The mystery box must contain a blue ball or a yellow ball. $\Box(\mathsf{A} \vee \mathsf{B})$
$\rightsquigarrow$ *The mystery box might contain a blue ball and it might contain a yellow ball*
a. $\Diamond\mathsf{A} \wedge \Diamond\mathsf{B}$
b. $\neg\Box\mathsf{A} \wedge \neg\Box\mathsf{B}$

Competing accounts have been proposed for the D-INFERENCES of disjunction under universal nominal quantifiers ([1, 2, 7, 10, 14]), but crucially, some of them predict disjunction under universal modals to give rise to D-INFERENCES independently of NU-INFERENCES ([14, 10, 1]), while others do not ([2, 3]).

To adjudicate between these opposite predictions, we carried out two sentence-picture verification experiments. The results replicate the earlier experimental finding that D-INFERENCES can be observed without NU-INFERENCES for disjunction under nominal universal quantifiers, but they also show that for disjunction under universal modals, D-INFERENCES can be observed independently, contrary to what has been assumed in the literature. In addition, we observe that D-INFERENCES trigger effects on response times that are the opposite of what is typically observed for scalar inferences, which we take to be a challenge for accounts of D-INFERENCES as a special case of scalar inferences.

2 Experiments

2.1 Participants

For each experiment, 100 native speakers of English were recruited online through Prolific.ac ([11]) using the same pre-screening criteria (first language: English, nationality: UK/US, country of birth: UK/US, prior approval rate: above 90%). Participants were paid £2.20, and average completion time was about 13 minutes. Participants gave written informed consent. Data were collected and stored in accordance with the provisions of Data Protection Act 2018.

2.2 Materials and Design

Examples of sentence-picture items used in Experiment 1 and Experiment 2 are given in Table 1. In each experiment, sentences were presented as being uttered by one of two characters, Sam or Mia, and the participants' task was to decide whether or not the character's utterance was a good description of the picture accompanying that utterance.

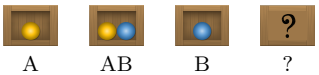
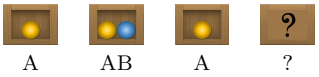
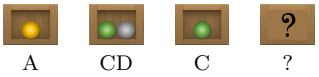
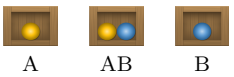
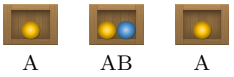
Example sentence	Example picture	Condition
Experiment 1		
NOMINAL *Every visible box contains a yellow ball or a blue ball.*	 A AB B ?	True
	 A AB A ?	Target-1
MODAL EPISTEMIC *The mystery box must contain a blue ball or a yellow ball.*	 A AC A ?	Target-2
	 A CD C ?	False
Experiment 2		
NOMINAL *Every box contains a yellow ball or a blue ball.*	 A AB B	True
	 A AB A	Target-1
MODAL DEONTIC *Sam must pick a box with a blue ball or a yellow ball.*	 A AC A	Target-2
	 A CD C	False

Table 1: Example items illustrating the experimental conditions for the target sentences in Experiments 1 and 2. There were 6 instances of each condition per quantifier type.

In Experiment 1, target sentences were NOMINAL sentences like (1), involving the determiner *every*, and MODAL sentences likes (4), in which *must* was used as an epistemic modal. Pictures involved three open ('visible') boxes, each of which contained one or two colored balls, and one closed ('mystery') box, with a question mark on it. Participants were instructed that the mystery box always has the same contents as one of the three visible boxes. They were asked to decide whether or not the sentence was a good description of what's inside the relevant box(es). In the control conditions, target sentences were paired with pictures that make them either TRUE or FALSE, regardless of the inferences under investigation. In the TARGET-1 conditions, they were paired with sentences that make their NU-INFERENCES false, but their D-INFERENCES true. In the TARGET-2 conditions, they were paired with sentences that make both these inferences false. Quantifier type (2 levels: NOMINAL, MODAL) and picture type (4 levels: TRUE, TARGET-1, TARGET-2, FALSE) were crossed to obtain a 2×4 within-subjects factorial design.

Experiment 2 was built on Experiment 1 by adapting the materials and instructions from Experiment 1 so as to extend our investigation to deontic modality. Target sentences were NOMINAL sentences with the determiner *every*, similar to those tested in Experiment 1, and novel MODAL sentences in which the modal *must* was used to express deontic necessity. Pictures were similar to those used in Experiment 1, except that they involved only three boxes, all open.

In the instructions, participants were told that the two characters were playing together and that, depending on the game, one of them either had to describe what's inside the boxes or had to pick one of the boxes. The first game scenario was used to test NOMINAL sentences and the second to test MODAL sentences. Depending on the game scenario, participants had to decide whether the utterance was a good description of the box(es) or of the character's options. The rest of the design of Experiment 2 was identical to that of Experiment 1 in all respects.

2.3 Procedure

The procedure was the same in both experiments, provided the differences in instructions we described above. NOMINAL trials and MODAL trials were presented in two separate blocks, the order of which was counterbalanced between participants. Each block started with a short training and included additional control items designed to check that subjects understood the instructions properly and had no problem understanding the quantifiers and connectives used in the target sentences. In each block, trials were presented in a random order. Participants provided their answers by clicking one of two response buttons, labelled 'Good' and 'Bad', respectively. Participants' responses and response times were recorded on each trial.

2.4 Data treatment

8 participants in Exp. 1 and 6 participants in Exp. 2 were excluded prior to analyses because their performance on the control items was below the pre-established threshold of 80% accuracy. For both experiments, participants' responses and response times on the target NOMINAL and MODAL trials were analysed using mixed-effects regression models with the maximal random effect structure justified by the design and supported by the data. Figure 1 shows the mean acceptance rate to the target sentences by quantifier type and experimental condition. Figure 2 shows the mean RTs by quantifier type, experimental condition and response type.

2.5 Responses

Overall, the response patterns for MODAL sentences were similar to those for NOMINAL sentences. In both experiments, both sentence types gave rise to intermediate acceptability ratings in their TARGET-2 conditions (comparisons with TRUE and FALSE: all $|\beta|$s$>$ 4.34, all ps$<$ .001), with a bimodal distribution of by-subject mean ratings (unimodality tests: all Ds$>$ 0.10, ps$<$ .001) while they were both uniformly accepted in their TARGET-1 conditions (comparisons with TRUE: all $|\beta|$s$<$ 0.42, ns; comparisons with FALSE: all $|\beta|$s$>$ 12.5, ps$<$ .001; unimodality: all Ds$<$ 0.05, ns). Responses to NOMINAL and MODAL sentences in the TARGET-1 and TARGET-2 conditions were further compared by testing the interaction between quantifier and picture types. A significant interaction was found in Experiment 1 ($\beta = -1,80$, $SE = .49$, $p < .001$), showing that the acceptability contrast between TARGET-1 and TARGET-2 was greater in the MODAL than in the NOMINAL cases; no such an interaction was found in Experiment 2 ($\beta = -0,42$, $SE = .32$, $p = 0.19$). These results replicate the main results from [7] and establish that the key contrasts previously observed for NOMINAL quantifiers reproduce with MODAL quantifiers.

2.6 Response times

In both experiments, NOMINAL and MODAL sentences showed response delay effects in the opposite direction to those commonly observed for regular scalar implicature in similar acceptability judgement tasks. Specifically, in the TARGET-2 conditions, participants took significantly longer

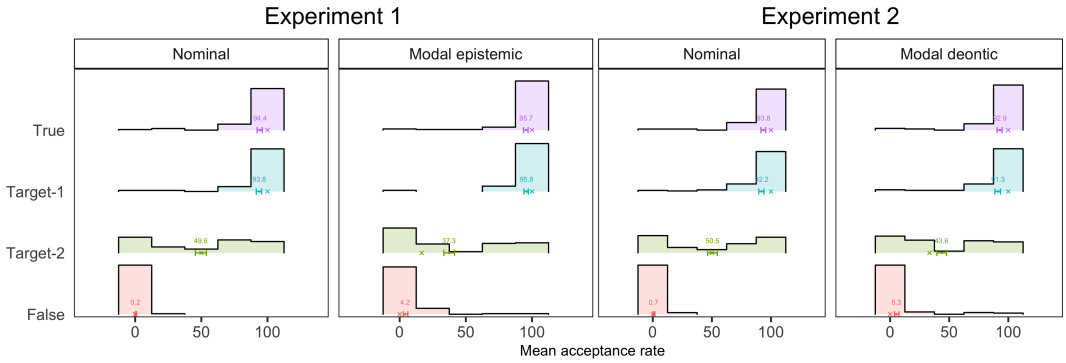


Figure 1: Mean acceptance rate by quantifier type and experimental condition. The distribution of by-participant mean ratings is visualised by a histogram, the grand mean by a thick bar with its value on top and the 95% CI around it, and the median by a cross.

to accept than reject both NOMINAL and MODAL sentences (all βs$>$ 333, all ps$<$ .01). By contrast, in the control conditions, participants were significantly faster to accept than reject these same sentences (all βs$<$ -241, all ps$<$ .001) while they were equally fast at rejecting them in their FALSE and TARGET-2 conditions in both experiments (all $|\beta|$s$<$ 38, ns) and equally fast at accepting them in their TRUE and TARGET-1 conditions in Experiment 1 (all $|\beta|$s$<$ 14, ns).

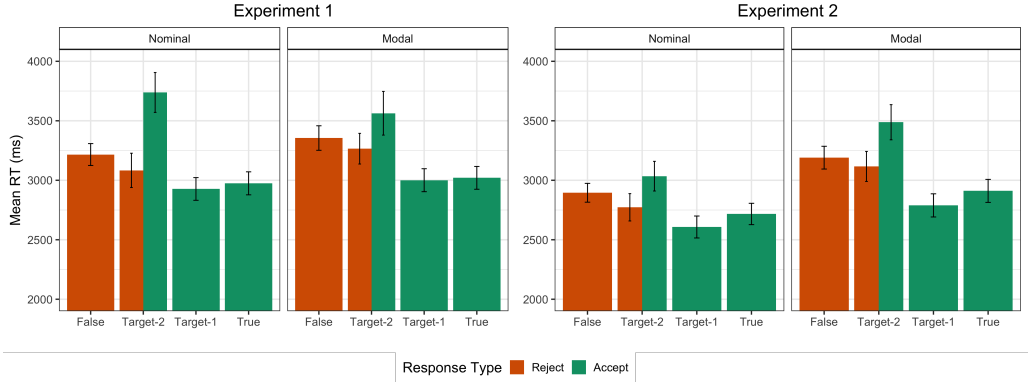


Figure 2: Mean RTs (in ms) by sentence type, experimental condition and response type. Error bars represent 95% CIs. RTs were analysed by considering correct responses in the control TRUE ('accept') and FALSE ('reject') conditions, 'accept' responses in the TARGET-1 conditions, and both 'accept' and 'reject' responses in the TARGET-2 conditions.

3 Discussion

We experimentally investigated the D-INFERENCES and NU-INFERENCES associated with disjunction embedded under the nominal universal quantifier *every* and the universal modal *must* in its epistemic and deontic readings. Our results show that disjunction under both *every* and *must* gives rise to D-INFERENCES independently of NU-INFERENCES. Also, the D-INFERENCES were overall observed to the same degree of robustness. The results for *every* essentially replicate previous finding by [7], but the results for *must* go against the claim in the literature

that for disjunction under a universal modal, the D-INFERENCE is always observed in tandem with the NU-INFERENCE ([3, 7, 9, 13]). Our results, therefore, present a challenge to accounts of D-INFERENCES which cannot derive them without NU-INFERENCES for disjunction under universal modals.[1]

In addition, we found response delay effects in the opposite direction to those generally observed for regular scalar implicatures in similar verification tasks ([4, 5] among others). Specifically, we found that it takes more time to accept than reject nominal and modal sentences with disjunction in cases where their D-INFERENCES and NU-INFERENCES are false, compared to cases where only their NU-INFERENCES are. This processing slowdown, established across multiple comparison points, suggests that, in contrast to what is usually observed with regular scalar inferences, it is the suspension rather than the derivation of the D-inferences that is costly. This finding is thus challenging for accounts of D-INFERENCES as a special case of scalar inferences.[2]

Taken together, the present findings suggest that we need a theory of D-INFERENCES that can derive them independently of NU-INFERENCES for both disjunction under universal nominal quantifiers and disjunction under universal modals. The reaction time data also suggest that they should not be treated as a species of scalar inferences, but as a different type of inference, e.g., as proposed by [1].

Lastly, we would like to remark on the discrepancy between our experimental finding and the introspective judgments reported in the literature for disjunction under universal modals. Most notably, [7] observes that (5) feels misleading in a context in which there is a requirement to wear sneakers in the gym (while running shorts are optional).

(5) You are required to wear sneakers or running shorts.

Importantly, the oddness of (5) is unexpected if the sentence readily receives a reading with a D-INFERENCE but without an NU-INFERENCE. We do not mean to deny the validity of the observation regarding (5). Rather, we take it as suggesting that D-INFERENCES cannot always be observed independently of NU-INFERENCES and that a theoretical account of D-INFERENCES ultimately must be able to explain what the relevant restrictions consist in and why these restrictions exist.

References

[1] Maria Aloni. Logic and conversation: the case of free choice. *Semantics and Pragmatics*, 15:5, 2022.

[2] Moshe Bar-Lev and Danny Fox. Free choice, simplification, and Innocent Inclusion. *Natural Language Semantics*, 28:175–223, 2020.

[3] Richard Jefferson Booth. Independent alternatives: Ross's puzzle and free choice. *Philosophical Studies*, 179(4):1241–1273, 2022.

[1] All such accounts can also derive a reading without either inference. This reading, however, is compatible with both our target conditions and, thus, it cannot account for the differences we found between the two.

[2] A similar challenge has been raised to implicature accounts of free choice inferences, which also do not give rise to the delay effects found with regular scalar implicatures ([6]). The main response to that challenge locates the source of this difference between free choice inferences and regular implicatures in the type of alternatives involved in their derivation. In particular, the idea is that the difference would be linked to the fact that regular implicatures are based on alternatives constructed by replacing lexical items in the asserted sentence, while free choice inferences only involve alternatives which are part of the asserted sentence. We note that this response does not extended straightforwardly to the case of D-INFERENCES. This is because, at least in some implicature accounts of D-INFERENCES (e.g. [2]), the alternatives involved in their derivation are constructed with lexical substitutions in the same way as in the case of regular implicatures.

[4] Lewis Bott, Todd M Bailey, and Daniel Grodner. Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language*, 66(1):123–142, 2012.

[5] Lewis Bott and Ira Noveck. Some utterances are underinformative. *Journal of Memory and Language*, 51(3):437–457, 2004.

[6] Emmanuel Chemla and Lewis Bott. Processing inferences at the semantics/pragmatics frontier: disjunctions and *free choice*. *Cognition*, 130(3):380–396, 2014.

[7] Luka Crnič, Emmanuel Chemla, and Danny Fox. Scalar implicatures of embedded disjunction. *Natural Language Semantics*, 23(4):271–305, 2015.

[8] Danny Fox. Free choice and the theory of scalar implicatures. In Uli Sauerland and Penka Stateva, editors, *Presupposition and Implicature in Compositional Semantics*, pages 71–120. Palgrave, 2007.

[9] Melissa Fusco. Deontic modality and the semantics of choice. *Philosophers' imprint*, 15, 2015.

[10] Serge Minor. Implicatures and events: A study of distributive inferences. *under review*, 2022.

[11] Stefan Palan and Christian Schitter. Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[12] Uli Sauerland. Scalar implicatures in complex sentences. *Linguistics and Philosophy*, 27(3):367–391, 2004.

[13] Geoffrey Sayre-McCord. Deontic logic and the priority of moral theory. *Noûs*, pages 179–197, 1986.

[14] Yasutada Sudo. Scalar implicatures with discourse referents: A case study on plurality inferences. *Linguistics and Philosophy*, to appear.

Exclusives vary in strength and scale structure: experimental evidence *

Eszter Ronai[1] and Lucas Fagen[2]

[1] Northwestern University, Evanston, Illinois, U.S.A.
ronai@northwestern.edu
[2] The University of Chicago, Chicago, Illinois, U.S.A.
lfagen@uchicago.edu

Abstract

This paper is an investigation of parameters of variation across English exclusive modifiers. We report on three experiments that test the robustness of exclusionary inference calculation (e.g., *merely intelligent* → *not brilliant*) across a large number of different lexical scales. Our findings reveal that 1) *just* excludes less robustly than *only*, 2) while *only* allows both complement-exclusion and rank-order readings, *merely* prefers rank-order ones, and 3) *just* and *only* are equally QUD-sensitive. We discuss these results in light of existing theoretical proposals about the semantics of exclusives.

1 Background

1.1 Exclusive modifiers

Exclusive modifiers, which in English include *only*, *just*, and *merely* (1), form a lexical class (Coppock and Beaver 2014), conveying that some proposition is true (the *prejacent*, (1-a)) and that alternatives to the prejacent are false (1-b).

(1) Mary **only/just/merely** ate the cookies.
   a. → Mary ate the cookies
   b. → Mary ate nothing other than the cookies

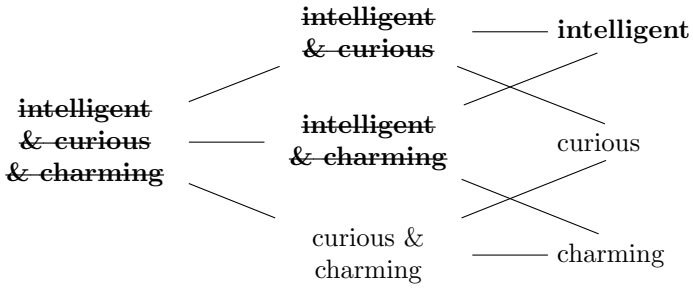
(2) The student is **only** intelligent.
   a. → The student is not curious, not charming, etc.
   b. → The student is not brilliant

Exclusives vary along different parameters. In this paper we focus on two: scale structure bias and strength of exclusion. It has been proposed that the excluded alternative set can vary in scale structure (Klinedinst 2005; Beaver and Clark 2008; Coppock and Beaver 2014). Specifically, exclusives can have 'complement-exclusion' readings (2-a) which exclude everything other than the prejacent, and 'rank-order' readings (2-b) which exclude alternatives ranked higher than the prejacent on a scale. These readings can receive a uniform semantic analysis, e.g. Coppock and Beaver (2014) analyze complement-exclusion and rank-order readings as excluding along differently structured scales (as shown in (3) and (4), adapted from their examples 27 and 29). But what we are primarily interested in in this paper is that exclusives do not always vary freely along this parameter: while *only* and *just* admit both readings, *merely* has been argued to prefer rank-order scales. Coppock and Beaver (2014) analyze variation in scale structure as resulting from "soft preferences" rather than absolute restrictions. Such effects are likely to

*We are grateful to three anonymous AC reviewers and the audience at the UChicago LEAP Workshop for their feedback. This material is based upon work supported by the National Science Foundation under Grant No. #BCS-2041312. Authors contributed equally to this work and are listed in reverse alphabetical order.

emerge more starkly across items in an experimental setting than via intuition alone, and are therefore worth testing directly.

(3)



(4)



The strength of the exclusion has also been argued to vary. *Just* but not *only* has a wider range of readings, paraphrasable with *simply*, that exclude alternatives understood as uninformative, unknown (5-a), redundant (5-b), or irrelevant, but not necessarily false. To explain this, various authors have proposed that *just* can operate on alternatives that *only* does not, including: covert causal modifiers with trivial semantic content (Wiegand 2018), answers to 'potential' questions in addition to the current QUD (Warstadt 2020), or metalinguistic alternatives at the speech act level (Beltrama 2022).

(5) a. The lights in this place **just**/**#only** turn off and on. (Warstadt 2020, ex. 1a)
Paraphrase: the lights turn off and on for no reason.

b. The pumpkin bisque is **just**/**#only** delicious! (Warstadt 2020, ex. 1b)
Paraphrase: the pumpkin bisque is extremely delicious.

Warstadt (2020) argues that such readings also require relaxing the truth-conditional status of the exclusive operation, proposing a distinction between "strong" exclusives which declare alternative propositions false, and "weak" exclusives which declare them unassertable. (See also Beltrama 2022, who proposes that emphatic *just* as in (5-b) declares alternatives not unassertable per se but not "worthy of assertion", p. 347.) Warstadt (2020) analyzes *only* as strong and *just* as weak. This is contra Coppock and Beaver (2014), who take both *only* and *just* to be strong, and have no explanation for weak readings with *just*. Warstadt's theoretical proposal regarding strength of exclusion is therefore novel and also worth testing experimentally.

1.2 Scalar diversity

Our testing ground in this paper is pairs of lexical items that form a scale. More specifically, we turn to the scalar diversity phenomenon: the observation that scalar expressions vary in how likely they are to lead to scalar implicature (SI) (i.a. van Tiel et al. 2016). A classic example of SI is (6): upon encountering an utterance containing *some*, hearers compute the negation of its stronger scalar alternative *all*. Similarly to ⟨*some*, *all*⟩, e.g. ⟨*intelligent*, *brilliant*⟩ also forms a scale: an utterance containing *intelligent* can lead to the SI *not brilliant* (7). But variation exists across different scales: the SI in (6) arises much more robustly than the one in (7).

(6) Mary ate some of the cookies.
→ SI: Mary ate some, but not all, of the cookies.

(7) The student is intelligent.
→ SI: The student is intelligent, but not brilliant.

Scalar diversity persists even in the presence of exclusives. Ronai and Xiang (2022) (henceforth R&X) found that even when sentences such as (6)-(7) contain *only*, variation still remains

in the likelihood of deriving a *not all* or *not brilliant* inference. This is puzzling, since while SI is a cancellable pragmatic inference, *only* encodes alternative exclusion in the semantics—which would predict uniformly ceiling-level inference rates. R&X hypothesized that interpretations of *only* were split between rank-order and complement-exclusion readings, leading to variation in whether the stronger scalar term was included in the alternative set. Given *The student is only intelligent*, the *not brilliant* inference would arise with rank-order *only*, but not necessarily with complement-exclusion *only*, which could be understood as excluding other unrelated properties (*not curious*, *not charming*, etc).

2 Experiments 1 and 2: *just* and *merely*

Though recent progress has been made in the theoretical literature describing variation among exclusives, much remains to be understood about which parameters vary and how. Here, we provide the first systematic experimental assessment of this domain, focusing on strength of exclusion and scale structure bias. First, we test Warstadt (2020)'s claim about strength of exclusion by comparing *just* vs. *only*. Second, we compare *only* vs. *merely*, the latter of which is claimed to prefer rank-order readings. This also addresses R&X's hypothesis that interpretations of *only* were split according to scale structure in their experiment.

2.1 Task and procedure

Experiment 1 was conducted on the web-based PCIbex platform (Zehr and Schwarz 2018). 40 speakers of American English were recruited on Prolific and were screened with a demographic survey and attention checks; data from 39 participants is reported below. The experiment used an inference task (i.a. van Tiel et al. 2016): participants saw sentences such as "Mary: *The student is just intelligent.*" and were asked the question "Would you conclude from this that Mary thinks the student is not brilliant?". They responded with "Yes", which indexes that the participant has calculated the exclusionary (*not brilliant*) inference, or "No", which suggests that the inference was not calculated. We used the same task and items as R&X's experiments, allowing for a direct comparison to their results. We tested 51 different lexical scales.

Experiment 2 was identical to Experiment 1 in its basic procedure, task and items, as well as participant recruitment and removal. Data from 35 participants is reported. Experiment 2 tested the exclusive *merely*. That is, participants saw stimuli such as "Mary: *The student is merely intelligent*".

2.2 Predictions

We make the following two predictions. First, given Warstadt (2020)'s claim that *just* is a weak exclusive, while *only* is a strong exclusive, we predict lower rates of inference calculation for Experiment 1 than was found for *only* (by R&X). Second, we predict higher rates of inference calculation for Experiment 2 than was found for *only*. This is because all our items test rank-order alternatives, and while *only* allows both complement-exclusion and rank-order readings, *merely* has been claimed to prefer rank-order readings (Coppock and Beaver 2014).

2.3 Results

For reasons of space, we present the visualizations of our experimental results in the Appendix. Figure 1 shows the results of both experiments, along with R&X's SI and *only* results. The

percent of exclusionary inference calculation corresponds to the percent of "Yes" responses in the inference task. To compare the rates of inference calculation in Experiments 1-2 to R&X's *only* experiment, we fit a logistic mixed effects regression model using the lme4 package in R (Bates et al. 2015). The model predicted Response ("Yes" vs. "No") as a function of Exclusive (*just* vs. *only* vs. *merely*). Random intercepts for participants and random slopes and intercepts for items were included. Since our predictions concern comparing *just* and *merely* to *only*, the predictor Exclusive was treatment coded, with *only* coded as the reference level. The model revealed significantly lower rates of inference calculation with *just* compared to *only* (Estimate=-0.7, SE=0.28, z=-2.5, p <0.05), as well as significantly higher rates of inference calculation with *merely* than was found with *only* (Estimate=0.96, SE=0.28, z=3.38, p <0.001). Averaged over the 51 different scales, the target exclusionary inference was calculated at the rate of 52.9% with *just*, 65.5% with *only*, and 80.2% with *merely*.

Since inference rates were lowest with *just*, the question may arise whether it can be maintained that *just* excludes alternatives semantically. To test this, we fit an additional statistical model comparing our Experiment 1 findings to R&X's SI experiment, which found an average rate of 33.1% SI calculation. The fixed effects predictor was sum-coded (SI: -0.5 and *just*: 0.5). SI rates were found to be significantly lower than the current Experiment 1 results with *just* (Estimate=1.32, SE=0.25, z=5.35, p <0.001). This confirms that alternative exclusion with all three exclusive modifiers is stronger than alternative exclusion via SI.

2.4 Discussion

Both predictions we made for Experiments 1-2 are borne out by the results. First, Experiment 1 found lower rates of exclusionary inference calculation with *just* than with *only*. This is consistent with the hypothesis that *just* excludes alternatives via a weaker semantic operation than *only* —a question we return to in Experiment 3. Second, Experiment 2 found higher inference rates with *merely* than with *only*. Since all our items test rank-order alternatives, this strongly supports the claim that *merely* biases toward rank-order scales (Coppock and Beaver 2014). These results also support R&X's hypothesis that participants in their experiment were split between rank-order and complement-exclusion readings of *only*. When a rank-order bias was introduced by *merely*, the stronger scalar term was more unambiguously understood as one of the salient alternatives, which led to an increase in calculating the target inference.

One may wonder whether different scales interact differently with the two tested parameters of variation. In order to check whether the relative order of different lexical scales remained consistent across manipulations, we calculated rank-order correlations using Kendall's τ_B, and found significant by-item correlations between experiments. Items with low SI rates continue to have relatively low inference rates even with (stronger) exclusives. As SI rates increase, so do inference rates with *just* (τ_B=0.59, p<0.001); as rates with *just* increase, so do rates with *only* (τ_B=0.59, p<0.001); and rates with *merely* are also correlated with *only* (τ_B=0.53, p<0.001). Only a small minority (≈5) of scales deviate from the general patterns. This highlights the role of lexico-semantic factors in the scalar diversity phenomenon.

3 Experiment 3: *just* + QUD

One explanation for our Experiment 1 results could be that *just* excludes alternatives via a weaker semantic operation than *only* (as suggested by Warstadt 2020, Beltrama 2022). If *just* were declaring scalar alternatives like *brilliant* uninformative or unassertable, rather than false, participants might have been more reluctant to answer "Yes" to questions like "Would you

conclude from this that Mary thinks the student is not brilliant?". Another explanation could be that *just* excludes a wider range of nonfocal alternatives. Consider Warstadt's proposal that *just* can answer potential questions ("intuitively possible future QUDs", p. 373) in addition to the current QUD. According to Warstadt, *just* in (8-a) signals that there are no assertable answers to the potential followup (8-b), "preventing the addressee from asking a useless question" (p. 373). On this account, *brilliant* or any other alternative to *intelligent* would not have been one of the excluded alternatives in our Experiment 1, because *just* was answering a potential question rather than the QUD itself.

(8) a. The lights in this place just turn off and on.
b. Why do the lights turn off and on?

A third possibility is that theories like Wiegand (2018)'s and Warstadt (2020)'s that aim to unify canonically exclusive and noncanonical readings of *just* under a single entry are on the wrong track. Instead, the lower rates of target inference calculation in Experiment 1 could reflect lexical ambiguity: participants excluded stronger scalar alternatives like *brilliant* when *just* was interpreted exclusively, and not otherwise.

To test these possibilities, we turn to another finding from the experimental pragmatics literature: explicit QUDs (Roberts 1996/2012) encourage inference calculation (i.a. Degen 2013; Zondervan, Meroni, and Gualmini 2008; Ronai and Xiang 2021; Ronai and Xiang 2022). For example, R&X found that when sentences with *only* were presented as answers to polar questions containing the stronger scalar term, participants were more likely to calculate the target exclusionary inference than in the null context *only* experiment. Given Warstadt's proposal that *just* can answer potential questions other than the current QUD, we would predict a differential effect of experimentally manipulating the QUD. Specifically, we would predict *just* to exhibit reduced sensitivity to the QUD compared to *only*.

3.1 Task and procedure

Experiment 3 had the same basic procedure, task, items, and participant recruitment and exclusion as the previous experiments. Data from 39 participants is reported. Similarly to Experiment 1, we tested the exclusive *just* in an inference task. But in this experiment, we embedded Mary's statements in a dialogue context, where another conversational participant, Sue, first asked a question. Sue's questions were polar questions that contained the stronger scalar term, while Mary's answers were slightly modified from Experiment 1 to ensure dialogue coherence (e.g., *the student* was changed to *he*). An example is shown below:

(9) Sue: Is the student brilliant?
Mary: He is just intelligent.

Participants again answered task questions like "Would you conclude from this that Mary thinks the student is not brilliant?" with either "Yes" or "No".

3.2 Predictions

To assess whether exclusives differ from each other in how sensitive they are to the QUD, we will compare the findings in Experiment 3 (*just* + QUD) to Experiment 1 (*just*) and R&X's *only* and *only* + QUD experiments. In their *only* + QUD experiment, experimental items were identical to (9), except Mary's answers contained *only*. We make the following predictions for the comparison of the four experiments. First, as we saw in Section 2, rates of calculating the target exclusionary inference should be higher with *only* than *just*. Second, since QUDs

generally encourage inference calculation, we predict higher rates for QUD experiments than for null context experiments. Lastly and most crucially, given Warstadt's proposal that *just* can answer non-QUD potential questions, we predict an interaction of exclusives and context, such that adding the QUD has less of an effect on *just* than on *only*.

3.3 Results

Figure 2 shows the results of Experiment 3 (*just* + QUD), along with Experiment 1 (*just*, repeated from Figure 1), as well as R&X's *only* (repeated from Figure 1) and *only* + QUD results. Averaged over the different scales, the target exclusionary inference was calculated at the rate of 78.7% in Experiment 3 —compare R&X's *only* + QUD experiment, which had a rate of 88.3%, as well as the null context exclusives from Section 2.3.

For the statistical analysis of the four experiments, we fit a logistic mixed effects regression model that predicted Response ("Yes" vs. "No") as a function of Context (QUD vs. null context), Exclusive (*just* vs. *only*) and their interaction. The maximal converging random effects structure (Barr et al. 2013) included random intercepts for participants and items, as well as random slopes for "Context" and "Exclusive" (but not their interaction) for items. The fixed effects predictors we sum-coded before analysis (null context: -0.5 and QUD: 0.5; *just*: -0.5 and *only*: 0.5). The analysis revealed a significant effect of Context, such that QUD experiments led to higher rates of calculating the target inference than null context experiments (Estimate=1.84, SE=0.25, z=7.39, p <0.001) and a significant effect of Exclusive, such that *only* experiments led to higher rates than *just* experiments (Estimate=0.86, SE=0.25, z=3.47, p <0.001). However, the interaction was not significant (Estimate=0.18, SE=0.46, z=0.39, p =0.7).

3.4 Discussion

We did not find the statistical interaction predicted from Warstadt's proposal, namely that *just* would be less sensitive to the QUD manipulation than *only*. We interpret this as speaking against a theory that proposes one unified semantics for exclusive and nonexclusive *just*, where this unified *just* can answer potential questions that are not the current QUD. Instead, we tentatively propose that our results are most compatible with there being several different lexical entries for *just*. It is possible that exclusive *just* answers the QUD, while other flavors of *just* are not exclusive and do not answer the QUD. Adding the explicit QUD could have increased the rate of QUD-sensitive *just* interpretations compared to Experiment 1, because participants assumed that Sue's question was relevant. This in turn led to a higher rate of exclusionary inference calculation, since the only flavor of *just* that answers the QUD is exclusive *just*. In this way a lexical ambiguity account would predict our finding that the QUD manipulation did raise inference rates for *just*, but that we did not find *just* to be less QUD-sensitive than *only*.

4 Conclusion

We presented novel experimental evidence testing variation across exclusive modifiers in English. Our results provided support for two claims from the theoretical literature: that exclusives vary in scale structure bias (with *merely* preferring rank-order readings) and in strength of exclusion (with *just* excluding less robustly). However, we did not find evidence for exclusives varying in their QUD-sensitivity, and in particular that *just* can answer questions that are not the current QUD. Our QUD results tentatively suggest that the apparent strength difference between *just* and *only* does not stem from variation in logical strength but instead from lexical ambiguity.

References

Barr, Dale J et al. (2013). "Random effects structure for confirmatory hypothesis testing: Keep it maximal". In: *Journal of Memory and Language* 68.3, pp. 255–278. DOI: 10.1016/j.jml.2012.11.001.

Bates, Douglas et al. (2015). "Fitting Linear Mixed-Effects Models Using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48. DOI: 10.18637/jss.v067.i01.

Beaver, David and Brady Clark (2008). *Sense and sensitivity: how focus determines meaning*. Chichester: Wiley-Blackwell.

Beltrama, Andrea (2022). "Just perfect, simply the best: an analysis of emphatic exclusion". In: *Linguistics and Philosophy* 45, pp. 321–364. DOI: 10.1007/s10988-021-09326-x.

Coppock, Elizabeth and David Beaver (2014). "Principles of the exclusive muddle". In: *Journal of Semantics* 31.3, pp. 371–432. DOI: 10.1093/jos/fft007.

Degen, Judith (2013). "Alternatives in Pragmatic Reasoning". PhD thesis. University of Rochester.

Klinedinst, Nathan (2005). "Scales and *only*". MA thesis. UCLA.

Roberts, Craige (1996/2012). "Information structure in discourse: Towards an integrated formal theory of pragmatics". In: *Semantics and Pragmatics* 5.6, pp. 1–69. DOI: 10.3765/sp.5.6.

Ronai, Eszter and Ming Xiang (2021). "Pragmatic inferences are QUD-sensitive: an experimental study". In: *Journal of Linguistics* 57.4, pp. 841–870. DOI: 10.1017/S0022226720000389.

— (2022). "Quantifying semantic and pragmatic effects on scalar diversity". In: *Proceedings of the Linguistic Society of America*. Vol. 7. 1, p. 5216. DOI: 10.3765/plsa.v7i1.5216.

van Tiel, Bob et al. (2016). "Scalar diversity". In: *Journal of Semantics* 33.1, pp. 137–175.

Warstadt, Alex (2020). ""Just" don't ask: exclusives and potential questions". In: *Proceedings of Sinn und Bedeutung 24*. Ed. by Michael Franke et al. Vol. 2, pp. 373–390.

Wiegand, Mia (2018). "Exclusive morphosemantics: *just* and covert quantification". In: *Proceedings of the 35th West Coast Conference on Formal Linguistics*. Ed. by Wm. G. Bennett, Lindsay Hracs, and Dennis Ryan Storoshenko. Somerville, MA, pp. 419–429.

Zehr, Jeremy and Florian Schwarz (2018). *PennController for Internet Based Experiments (IBEX)*. https://doi.org/10.17605/OSF.IO/MD832.

Zondervan, Arjen, Luisa Meroni, and Andrea Gualmini (2008). "Experiments on the Role of the Question Under Discussion for Ambiguity Resolution and Implicature Computation in Adults". In: *Proceedings of Semantics and Linguistic Theory (SALT) 18*. Ed. by Tova Friedman and Satoshi Ito, pp. 765–777.

Appendix

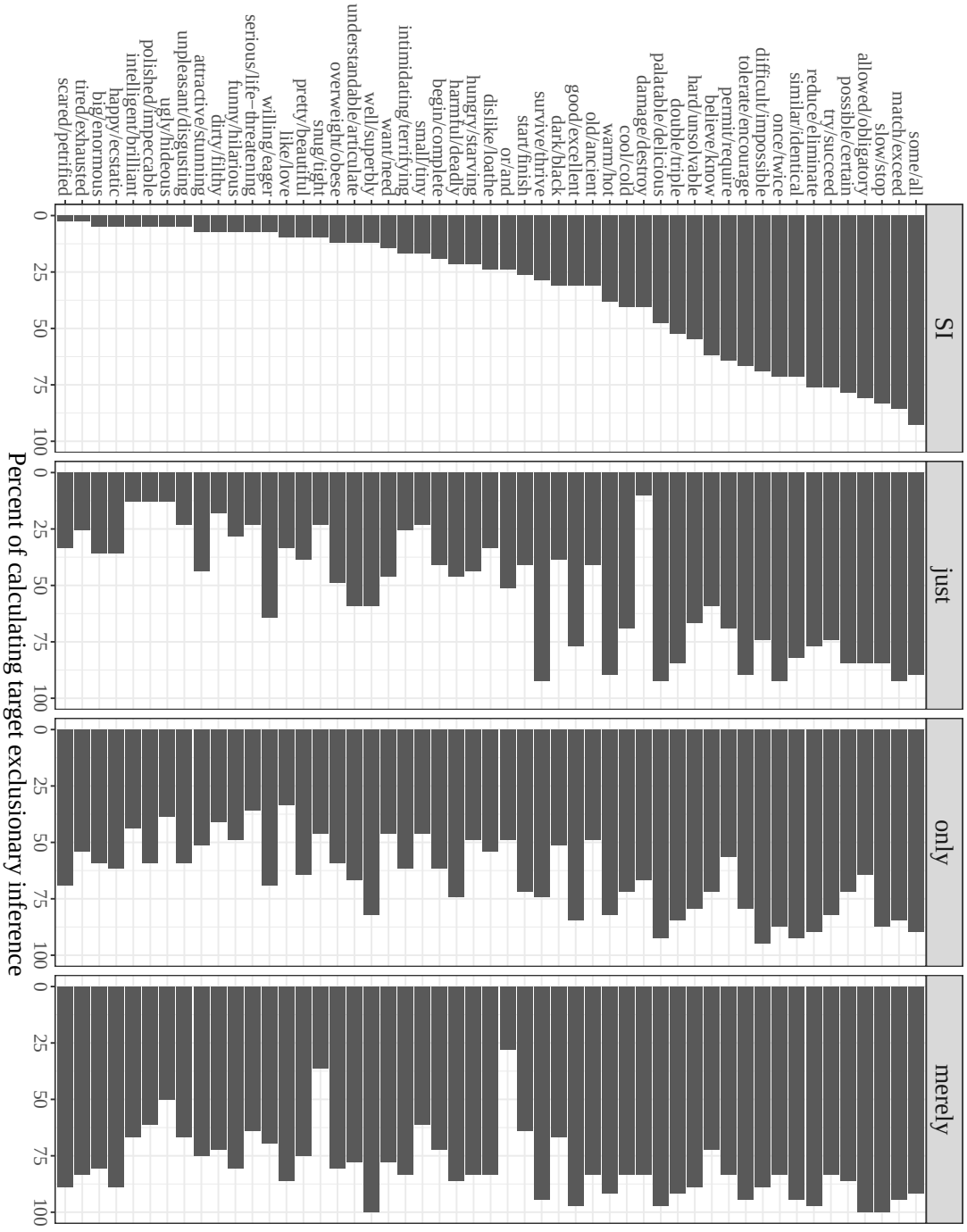


Figure 1: Results of Experiment 1 (*just*), Experiment 2 (*merely*), and Ronai and Xiang (2022)'s Experiment 1 (SI) and Experiment 3 (*only*).

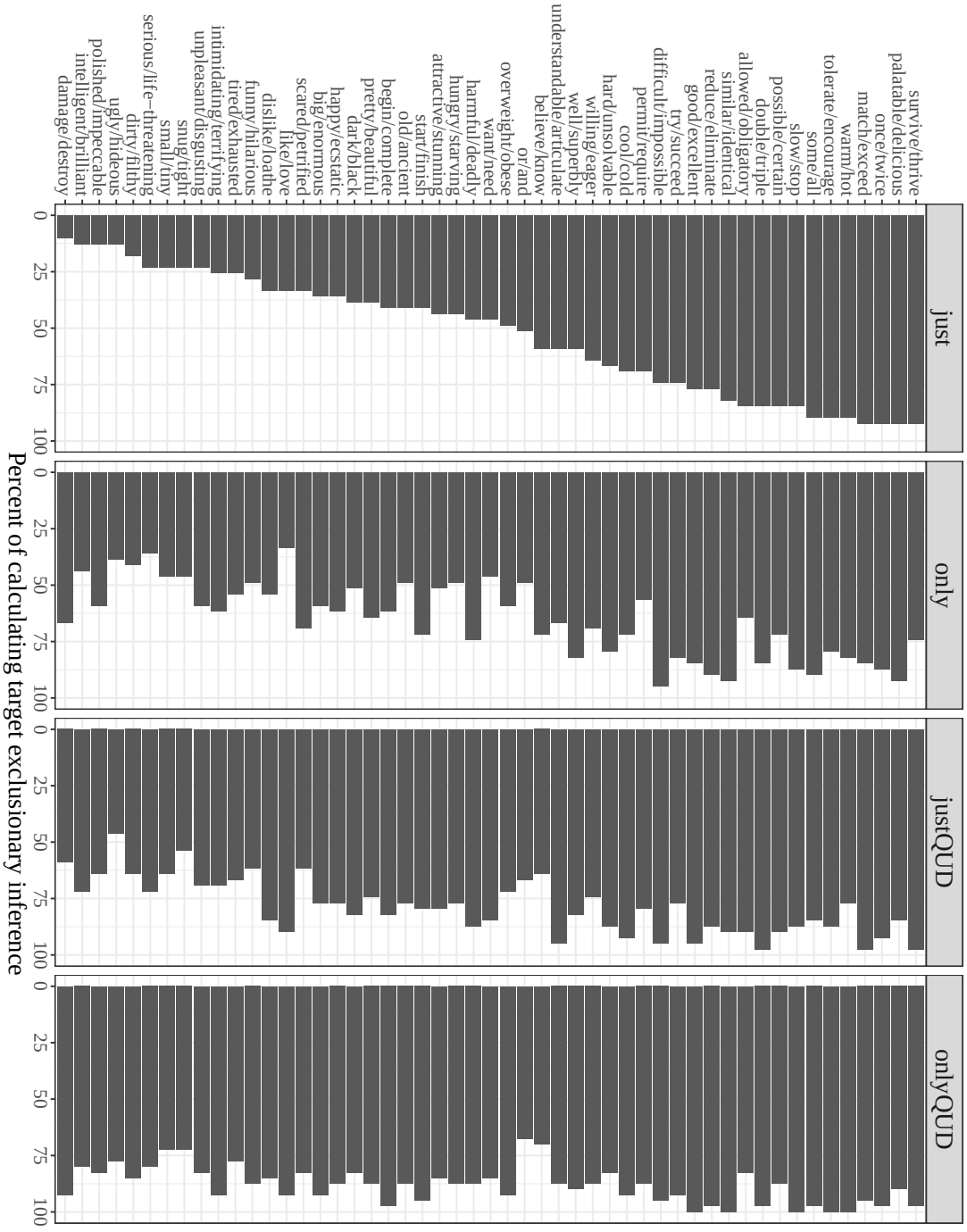


Figure 2: Results of Experiment 1 (*just*), Experiment 3 (*just* + QUD), and Ronai and Xiang (2022)'s Experiment 3 (*only*) and Experiment 4 (*only* + QUD).

Ability Modals as Causal Modals*

Paolo Santorio

University of Maryland, College Park
College Park, Maryland, USA

Abstract

Can and *be able to* can be used to describe what is possible in light of a subject's abilities. These modals give rise to several logical anomalies, which have led to analyses that stray far from standard Kratzer-style treatments (see [2], [10], [19] a.o.). I suggest that the source of the anomalies is a presupposition sitting on top of a standard existential semantics. The resulting account is both more conservative and more successful at accommodating the data than the alternatives. The presupposition of *S can/is able to* A states (roughly) that a state of S is causally sufficient for A. Interestingly, this account diverges from many existing account in that it does not invoke a notion of agency.

1 Introduction

Can and *be able to* can be used to describe what is possible in light of someone's abilities. Modals of this flavor (which I call 'modals of ability', or MoA for short) give rise to several logical anomalies. These anomalies have led to analyses that stray far from standard Kratzer-style treatments: in particular, they have been analyzed as involving a metalanguage conditional (see [19], a.o.) or a double layer of modal quantification, in which a possibility modal scopes over a necessity modal (see [2], [10], a.o.).

In this paper, I have two goals. First, after reviewing some of the logical anomalies generated by MoA, I argue that these anomalies mark the presence of a presupposition. Second, I suggest that the relevant presupposition states the existence of a causal connection between the subject of the modal and the event picked out by the prejacent. Specifically, the presupposition requires that a state of the subject is causally sufficient (in a way to be made precise) for the prejacent. A Kratzer-style existential semantics, supplemented with this presupposition, captures all the data. Notably, this analysis divorces MoA from a notion of agency, contrary to many accounts in the literature.

I proceed as follows: I describe some of the main puzzles about MoA in §2, give some motivation for a causal account in §3, and develop my analysis in §4.

2 Modals of ability and logical anomalies

2.1 The target of the analysis

My target are occurrences of *can* and *be able to* that track ability, as in (1a) and (1b).

(1) a. Ava can/is able to hit the bull's eye on the next throw.
b. Ben can/is able to join the conference virtually.

*Thanks to Harjit Bhogal, David Boylan, Fabrizio Cariani, Ilaria Canavotto, Matt Mandelkern, Dilip Ninan, Alexander Williams, and audiences at the University of Maryland and the Dublin Language Workshop.

I assume that, in the sentences in (1), *can* and *be able to* express what is possible in light of the abilities of a subject. Hence, in line with the literature on MoA, I assume that there is a dedicated modal flavor that tracks ability.Moreover, I assume that *can* may be used to express a variety of modal flavors (ability, circumstantial, dentic), but *be able to* lexicalizes the ability flavor. In support of this assumption, notice that (2a) can be uttered in a situation where Ava is a first time dart thrower (and hence, presumably, the *can* occurring in it expresses a non-ability modal flavor). But (2b) can't.

(2) a. Ava can hit the bull's eye on this throw—sometimes beginners are lucky!
b. # Ava is able to hit the bull's eye on this throw—sometimes beginners are lucky!

Again in line with the literature, I also assume that MoA have **specific** and **generic** occurrences. Specific occurrences, exemplified by (3a), describe a specific event that is possible in the light of the subject's abilities. Generic occurrences, exemplified by (3b), describe what is possible in general for a subject, given their abilities.

(3) a. Ava is able to hit the bull's eye on this next throw.
b. Ava is able to hit the bull's eye when playing darts.

I assume that specific uses are basic and I will focus on them throughout the paper.

2.2 Logical anomalies

On a classical Kratzer-style analysis ([14], [15], [16]), MoA are treated as existential modals with a realistic modal base (for all w, $w \in f(w)$). The schematic truth conditions of an ability report are:

(4) $[\![S \text{ can/is able to } \mathsf{A}]\!]^{w,f} = 1$ iff $\exists w' \in f(w)$ such that $[\![S \ \mathsf{As}]\!]^{w',f} = 1$

This simple analysis is challenged by several logical puzzles. Here I rehearse two.

Puzzle 1: variable strength. MoA appear to be stronger than (4) in positive contexts, but exactly as strong as (4) under negation.

The argument for strength in positive contexts comes in two steps. First, MoA are stronger than standard circumstantial modals. Consider (5) (adapted from [19]):

Jim and Jo are playing darts; Jo's child Susie, who is quite bad at darts, exclaims:

(5) # Let me take the shot! I can/am able to hit the bull's-eye on this throw.

(5) seems not true, even though there is a circumstantial possibility where Susie hits the bull's-eye out of sheer luck. Second, in positive contexts MoA are stronger in a further way: the truth of *S As/will A* is insufficient for the truth of *S is able to A*. Several authors have produced arguments for this point ([13], [1][1]). Here is a quick and simple one. Suppose that *S As/will A* does entail *S is able to A*. Then, since (6a) entails (6b), it also entails (6c).

(6) a. Maria will hit the target out of sheer luck on this throw.

[1] Boylan's position comes with a twist: he argues that ⌜*S* As⌝ does not entail ⌜*S can A*⌝ when A has a present or a future reference time, but that it does when the A has a past reference time. I don't have space here to engage with Boylan's position. Let me just say that I think that the data Boylan exploits is complicated by the interactions between modals and aspect: in particular, all root modals give rise to so-called actuality entailments (see [7]) when they combine with perfective aspect. This might have a crucial effect on the judgments.

b. Maria will hit the target on this throw.

c. Maria is able to hit the target on this throw.

From this, given minimal assumptions about conditionals, it follows that (7) should be a logical truth. But (7) is not a logical truth, and in fact it sounds pretty awkward.[2]

(7) If Maria will hit the target out of sheer luck on this throw, Maria is able to hit the target on this throw.

At the same time, MoA appear as weak as (4) predicts in negative contexts. Consider again Susie's case, and suppose that Susie's mom, Jo, responds to her by saying:

(8) No, you're not able to/you can't hit the bull's eye on this throw. You're not good enough.

(8) is the negation of Susie's original claim (5). And like the latter, it seems defective. Susie might get lucky and hit the bull's eye just on this throw. This seems enough to make (8) unassertable.

A second mark of weakness is that the truth of *S cannot A* seems sufficient for the truth of *S does not A/won't A*. To see this, notice that conjunctions like (9) are contradictory:

(9) # Susie is unable to hit the dartboard on this throw, but she will hit the dartboard on this throw (by sheer luck).

In sum, MoA appear to vindicate the following two principles.

Actuality does not entail ability.	$A \nvDash \Diamond A$	*S* will A $\nvDash$ *S* is able to A
Inability entails inactuality.	$\neg\Diamond A \vDash \neg A$	*S* is unable to A $\vDash$ *S* won't A

This combination is puzzling, since the two principles are contrapositives of each other.

Puzzle 2: failure of DoD. Kenny ([13]) famously presents the following case:

> **Context.** Clem is a somewhat experienced dart player. She's good enough to reliably hit the board when she throws, but she is still unable to control what section of the board she hits.

(10) a. Clem is able to hit the top or the bottom part of the board. ✓

b. Clem is able to hit the top part of the board. ✗

c. Clem is able to hit the bottom part of the board. ✗

These judgments violate Distribution over Disjunction, which is valid on standard semantics:

Failure of Distribution over Disjunction (DoD). $\Diamond(A \vee B) \nvDash \Diamond A \vee \Diamond B$

2.3 The mark of presupposition

The first puzzle suggests that MoA involve a presupposition of sorts. The reason is that notions of entailment that track presupposition, like Strawson entailment ([5]), don't contrapose (see [3] for related discussion). To see this, notice that the inference in (11a) is valid on a notion of Strawson-entailment, but its contrapositive in (11b) clearly is not.

(11) a. Ben doesn't know that Clem won. $\vDash$ Clem won.

b. Clem didn't win. $\nvDash$ Ben knows that Clem won.

[2]The argument relies on the following assumptions: (i) conditionals of the form *If A, then A* are logical truths, and (ii) if B^+ entails B, *If A, then* B^+ entails *If A, then B*. Both are fairly uncontroversial.

3 Evidence for a causal analysis

A further feature of MoA is that their complement seems subject to a 'no-chanciness' requirement. Suppose that Ava has bought a ticket for a fair lottery, which she has a genuine chance to win. Both sentences in (12) are infelicitous.

(12) a. ?? Ava is able to win the lottery.
b. ?? Ava is unable to win the lottery.

To accommodate cases like (12) (and, sometimes, as part of an account of DoD failure), several analyses ([19], [6], [25] a.o.) link MoA to a notion of agency. Roughly, the idea is that MoA track possibilities that an agent may bring about by acting in one of several ways. If this was on the right track, it might be used to rule out cases like (12a) and (12b), since winning the lottery (or losing it) is not an action, and not something an agent has control over.

As I argue at greater length in [23], however, agency is way too restrictive to account for the full range of uses of MoA in natural language. There are occurrences of MoA in which the subject is not an agent and/or the complement does not describe an action. Below are some natural examples of this sort.

(13) a. Steel is able to withstand a pressure of 100 tons.
b. This type of coal is able to burn without producing smoke.

(14) a. Ava is able to process sugar very fast now (since she's on a special medication).
b. Ben is able to perceive unusually high frequencies now (since he's on drugs).
c. Clem is able to fight off the virus over the next days.

I suggest that the analysis of MoA should invoke causal rather than agential notions.

4 The Causal Analysis

The causal notion that plays a key role in my analysis is **causal sufficiency**. Here I show how this notion can be defined precisely appealing to the causal models framework ([8], [21], [9], a.o.). I use causal models merely because they are a widespread and fruitful way of capturing causal notions formally; in principle, alternative treatments are possible.

4.1 Defining causal sufficiency

The idea I want to capture is this. Suppose that an expert dart thrower (say, Ava) and an inexpert one (say, Ben) are both about to take a throw. Intuitively, Ava's intention to throw, given appropriate background circumstances (no sudden gusts of wind, no one pushes her, etc) is sufficient to bring about that the target is hit. Not so for Ben. Whether Ben hits the target depends also on other factors, (which, for simplicity, we might identify with chance).

To characterize precisely the relevant notion of causal sufficiency, I use the causal models framework. Causal models are formal models used to charaterize causal processes. Here I understand a causal model as consisting of three elements: a set of **random variables**, a set of **structural equations**, and an **assignment of values** to the variables.

A **random variable** can be thought of as a set of mutually exclusive and jointly exhaustive outcomes for a process. Following standard notation, I represent random variables with uppercase Roman letters, like 'X'. I represent the values of random variables with lowercase Roman letters, sometimes with subscripts, like 'x_i'. Moreover, to say that variable X has value x_i, I

use the notation '$X = x_i$'. Notice that an equation like '$X = x_i$' basically says that a certain event obtains. (For example, using variables from the model below, we can represent the event of a target being hit with the equation '$T = 1$'.)

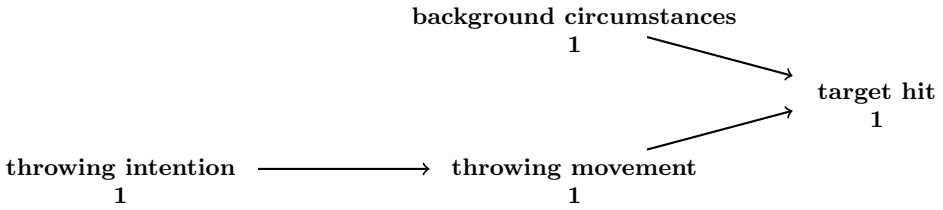
Structural equations are mathematical equations that state the relations between different values of random variables. An **assignment of values** is simply a mapping of each variable in a model to one of its possible values.

Let me go through an example. We can represent the case of an expert dart thrower (say, again, Ava) who is about to throw a dart with the following model:

Random variables			**Structural equations**
C	:	whether the background circumstances are favorable	
I	:	whether Ava has the intention to throw	$M = I$
M	:	whether Ava performs a correct throwing movement	$T = \min(C, M)$
T	:	whether Ava hits the target	

To complete our model, suppose also that all the variables have value 1 (corresponding to 'true') and that hence the target is hit.

Causal models can be can represented via graphs. Here is the graph for our toy model:



Now, against this background, we can define causal sufficiency as follows:

> **Causal sufficiency.** A set of events $\{X_1 = x_1, \ldots X_n = x_n\}$ is **causally sufficient** for an event $E = e$, relative to a set of equations $\mathcal{S}$ and background circumstances $\mathcal{C}$ iff any model that is (i) consistent with $\mathcal{S}$, (ii) consistent with $\mathcal{C}$, and (iii) such that $\{X_1 = x_1, \ldots X_n = x_n\}$, is also a model where $E = e$.

The idea behind the definition is simple. A set of events $\mathcal{E}$ is causally sufficient for an effect e just in case, holding fixed the equations and background circumstances, any model where the events in $\mathcal{E}$ obtain is also a model where e obtains.[3]

It's easy to check that, on this definition, Ava's throwing intention is causally sufficient for the target being hit, given that favorable background circumstances obtain. I leave it to the reader to verify that the sufficiency condition is not satisfied in a case involving an inexperienced thrower like Ben (where variables other than I contribute to determining the value of M).

4.2 Semantics

I suggest that the semantics of *able* involves reference to a notion of causal sufficiency. Specifically, *S is able to/can* A requires that there is a state of S that, given background circumstances, is causally sufficient for A. For reasons of space, I won't expand on what counts as a 'state' of the relevant sort. What states are relevant is clear enough in all the cases discussed in this paper. For example, in the case of the expert thrower Ava, we can take the relevant state to be the obtaining of a throwing intention ($I = 1$).[4]

[3] In passing, let me notice that this definition is reminiscent of the so-called 'INUS' account of causation, first introduced by [18] and then elaborated by many authors. See [20] for discussion.

[4] In general, states can be identified with sets of events in a causal model.

In the lexical entry of the modal, the causal requirement is encoded in a presupposition. Specifically, I propose that *can* and *be able to* work as existential quantifiers over a set of circumstantial worlds. On top of this existential meaning, a presupposition captures the causal requirement. The content of the presupposition is, roughly:[5]

> Either not-A, or there is a state of S that, given appropriate background circumstances, is causally sufficient for A

I adopt a fairly standard Kratzer-style framework, with one addition. On Kratzer's semantics, the interpretation of modals is relativized to a modal base and an ordering source, both of which are functions from worlds to sets of propositions. I add a third parameter, which specifies a function from worlds to sets of 'background propositions' that are used for assessing causal sufficiency.[6] Following Kratzer, I use 'f' to denote the modal base and 'B' to denote the background parameter. For simplicity, I ignore the ordering source. Here is the semantics:

(15) $[\![S \text{ can/is able to } \mathsf{A}]\!]^{w,f,B} =$

true iff $\exists w' \in \text{CIRCUMSTANTIAL}_{w,f,B}$ such that $[\![S \text{ As}]\!]^{w',f,B} = 1$

defined iff $\forall w' \in \text{CIRCUMSTANTIAL}_{w,f,B}$, either $[\![S \text{ As}]\!]^{w',f,B} = 0$, or $[\![S \text{ As}]\!]^{w',f,B} = 1$ and, in w', there is a state of S that is causally sufficient for A given $B(w)$.

Let me briefly sketch how this semantics accounts for the puzzles in §2.

The strength puzzle. In positive contexts, the presupposition requires (via the right disjunct) that, in every circumstantially accessible world where S As, there is a state of S that is causally sufficient for A. Hence an utterance of *S is able to A* requires that there is a circumstantially accessible world where S As, and in this world some state of S is causally sufficient for A. Conversely, in negative contexts the presupposition is immediately satisfied via the truth conditions (via the left disjunct), so it generates no extra requirements.

DoD failure. Consider again (10a). In a scenario where Clem is about to throw, there is a circumstantial possibility where there is a state of of Clem's (say, her having the intention to throw) that is causally sufficient for her to hit the board. But there are no circumstantial possibilities where Clem's being in that state is causally sufficient for her to hit the top part of the board, or causally sufficient for her to hit the bottom part of the board. Hence (10a) is true, but (10b) and (10c) are undefined.

5 Conclusion

Modals of abilty give rise to several logical anomalies, which have led to analyses that stray far from standard Kratzer-style treatments. In this paper, I have made two suggestions. (i) The source of the anomalies is a presupposition sitting on top of a standard existential semantics. (ii) Roughly, the presupposition requires that there is a state of the subject that is causally sufficient for the prejacent. I have shown how the relevant notion of causal sufficiency can be made precise by appeal to causal models. Causal notions have already turned out to be useful in modal semantics, notably in the semantics of counterfactuals (see [24], [12], [11], [4], [22], a.o.). It is unsurprising that they can also be useful for other flavors of modality.

[5] This presupposition resembles the so-called homogeneity requirement; see [17] for discussion.

[6] As a default, I assume that, for all w, $B(w) \subseteq f(w)$, i.e. the 'background propositions' are a subset of the propositions in the modal base.

References

[1] David Boylan. Does success entail ability? *Noûs*, 56(3):570–601, 2022.

[2] Mark A. Brown. On the logic of ability. *Journal of Philosophical Logic*, 17(1):1–26, 1988.

[3] Fabrizio Cariani and Simon Goldstein. Conditional heresies. *Philosophy and Phenomenological Research*, (2):251–282, 2018.

[4] Ivano Ciardelli, Linmin Zhang, and Lucas Champollion. Two switches in the theory of counterfactuals: A study of truth conditionality and minimal change. *Linguistic and Philosophy*, 41:577–621, 2018.

[5] Kai von Fintel. Npi licensing, strawson entailment, and context dependency. *Journal of Semantics*, 16(2):97–148, 1999.

[6] Melissa Fusco. Agential free choice. *Journal of Philosophical Logic*, 50(1):57–87, 2021.

[7] Valentine Hacquard. *Actuality Entailments*, pages 1–26. John Wiley & Sons, Ltd, 2020.

[8] Joseph Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

[9] Joseph Halpern. *Actual Causality*. MIT Press, 2016.

[10] John F. Horty. *Agency and Deontic Logic*. Oxford, England: Oxford University Press, 2001.

[11] Thomas Icard. From programs to causal models. In *Proceedings of the 21st Amsterdam Colloquium*, pages 35–44, University of Amsterdam, 2017.

[12] Stefan Kaufmann. Causal premise semantics. *Cognitive science*, 37(6):1136–1170, 2013.

[13] Anthony Kenny. Human abilities and dynamic modalities. In *Essays on explanation and understanding*, pages 209–232. Dordrecht: Reidel, doi=https://doi.org/10.1007/978-94-010-1823-$4_1$1, 1976.

[14] Angelika Kratzer. What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1(3):337–355, 1977.

[15] Angelika Kratzer. The notional category of modality. In H. J. Eikmeyer and H. Rieser, editors, *Words, Worlds, and Contexts: New Approaches to Word Semantics*. de Gruyter, Berlin, 1981.

[16] Angelika Kratzer. *Modals and Conditionals: New and Revised Perspectives*, volume 36. Oxford University Press, 2012.

[17] Manuel Križ. *Aspects of homogeneity in the semantics of natural language*. PhD thesis, University of Vienna, 2015.

[18] J.L. Mackie. Causes and conditions. *American Philosophical Quarterly*, 2(4):245—-264, 1965.

[19] Matthew Mandelkern, Ginger Schultheis, and David Boylan. Agentive modals. *Philosophical Review*, 126(3):301–343, 2017.

[20] L. A. Paul and Ned Hall. *Causation: A User's Guide*. Oxford University Press UK, 2013.

[21] Judea Pearl. *Causality: models, reasoning and inference*. Cambridge University Press, 2000.

[22] Paolo Santorio. Interventions in premise semantics. *Philosophers' Imprint*, 19(1), 2019.

[23] Paolo Santorio. Ability as causal modality. draft, University of Maryland, College Park, 2022.

[24] Katrin Schulz. If you'd wiggled A, then B would've changed: Causality and counterfactual conditionals. *Synthese*, 179(2):239–251, 2011.

[25] Malte Willer. Two puzzles about ability can. *Linguistics and Philosophy*, 44(3):551–586, 2021.

Breaking the fourth wall and (meta)fictional reference *

Merel Semeijn[1]

Institut Jean Nicod (École normale supérieure), Paris, France
semeijn.merel@gmail.com

Abstract

I investigate statements in fiction that 'break the fourth wall' (i.e., statements through which a fictional character somehow acknowledges the fictionality of their world) and suggest that they are a mirror image of 'parafictional statements', i.e., statements about what is true in some fiction. I explore two possible analyses, according to which statements that break the fourth wall are either a type of fictional statement or are a type of metafictional statement, and propose a synthesis of these two analyses.

We talk about fictions and fictional entities in a variety of meaningful ways. Following Recanati's [11] terminology, we can distinguish between at least three different kinds of discourse that involve reference to fictional entities: 'fictional', 'metafictional' and 'parafictional'. Consider the following statements:

(1) Frodo had a very trying time that afternoon

(2) Frodo was invented by Tolkien

(3) Frodo is a fictional character

(4) In *The Lord of the Rings*, Frodo lives in the Shire

(1) is a direct quote from Tolkien's *The Lord of the Rings* and is a fictional statement, i.e., a statement that is part of a fiction. Fictional statements are often analysed as pretend-assertions that are merely pretend-true. Fictional names (i.e., names of fictional entities) in fictional statements are likewise analysed as pretend-referring to actual objects. For instance, the name 'Frodo' in (1) involves pretend-reference to a flesh and blood hobbit called Frodo and (1) is only pretend-true (rather than really true). (2) and (3) are metafictional statements. Metafictional statements are statements about fictional entities *as fictional entities*. Metafictional statements are standardly analysed as regular assertions that are really true or false. Fictional names in metafictional statements are often analysed as referring to actually existing abstract objects. For instance, (2) is a true assertion because it is really true (not merely pretend-true) of the fictional character (or abstract object) Frodo that it was invented by Tolkien.

(4) is a parafictional statement, i.e., a report about the content of some fiction, i.e., a report on what is true in or according to some fiction.[1] As recently described by Recanati [11], the semantic analysis of parafictional statements raises some interesting questions. One option is to analyse them as a species of fictional statements (e.g. Evans [4]): 'Frodo' in (4) pretend-refers to a flesh and blood individual and (4) is a continuation of the pretence initiated by Tolkien and hence merely pretend-true. However, this is not entirely intuitive. It seems that parafictional statements are *really* true or false, depending on what the world is like. Given how Tolkien wrote his books, (4) is *actually* true (not just pretend-true) and its negation is *actually* false. Alternatively, we may analyse parafictional statements as a species of metafictional statements

*Many thanks to three anonymous Amsterdam Colloquium reviewers and to the audiences at the Milan MELT seminar, PLM6 and the *Fiction & Narrative across Media* workshop.

[1]This is how theorists usually characterize parafictional discourse. I have argued elsewhere [10] that we need to distinguish between truth *in* and truth *according to* some fiction.

(e.g. Zalta [15] or Inwagen [14]): Parafictional statements such as (4) are a type of assertion and hence can be really true or false. However, this kind of analysis seems to lead to the so-called 'problem of the wrong kind of object' (see Klauk [7]): On this analysis, 'Frodo' in (4) refers to an abstract object, so doesn't it then imply that in *The Lord of the Rings*, some abstract object lives in the Shire? Surely, that cannot be right either. An abstract object is not the right kind of object to live in some region.[2]

The following table from Rouille (p.c.) may shed light on why the analysis of parafictional statements is challenging:

	Genuine assertion	*Pretend assertion*
Reference to a flesh and blood individual	Parafictional	Fictional
Reference to an abstract object	Metafictional	?

Fictional statements (e.g., (1)) are classified as being made from a perspective internal to the fiction (they are pretend assertions and only pretend-true) and fictional names in them standardly involve internal pretend-reference to flesh and blood individuals. Metafictional statements (e.g, (2)) are made from a perspective external to the fiction (they are genuine assertions and really true or false) and fictional names in them involve real reference to abstract objects. Parafictional statements (e.g., (4)) at least *prima facie* seem to mix perspectives that are internal and external to the fiction: They are made from a perspective external to the fiction (they are genuine assertions that can be really true or false) but fictional names in them standardly involve internal pretend-reference to flesh and blood individuals, rather than to abstract objects.

This paper investigates what type of statement could be placed in the blank cell that mirrors the parafictional cell, i.e., statements that are (at least *prima facie*) internal to the fiction but involve metafictional external reference. I suggest that statements in fiction that 'break the fourth wall' (BtFW-statements), i.e., statements through which a fictional character somehow acknowledges their own fictionality, are such a mirror case. After a brief introduction of the phenomenon of breaking the fourth wall (section 1), I will explore two possible analyses, according to which BtFW-statements are either a type of fictional statement or are a type of metafictional statement, and propose a definition of BtFW-statements which is a synthesis of these two analyses (section 2).

1 Breaking the Fourth Wall

The concept of a 'fourth wall' originates in Diderot's [1] writings on theatre. On a traditional stage, if characters are in a room, then it is true in the fiction that they are surrounded by four walls. These walls are solid for the fictional characters on stage. However, one of these walls – the fourth wall – is transparent for the audience. Traditional theatre convention dictates that the theatre audience can 'peer into the fictional world' through this fourth wall whereas the fictional characters are oblivious to this audience. Generalising this concept to other media, the term 'fourth wall' denotes the ontological barrier or separation between the fictional and the actual world. In film, the fourth wall will often be a point of view that can move around in the fictional world and coincides with the location of the cameras that were used during filming. The question of where the fourth wall is to be located in other kinds of media (e.g. novels, pretend play/LARPing or videogames[3]) can become complicated. However, although it may be

[2] See Semeijn and Zalta [12] for a recent defence of the metafictional analysis against this objection.

[3] Some game scholars have argued that the inherent interactivity of videogames entails that there is no fourth wall in the case of video games. See Mosselaer [13] for an overview and the opposite view.

difficult to physically locate it, I assume that for any fiction (across different media) there is in fact an ontological divide (a 'fourth wall') between the fictional world and actual world.

This ontological barrier can seemingly be 'broken'. The common pretheoretical and intuitive understanding of the phenomenon of breaking the fourth wall is that of a fictional character acknowledging that there is such an ontological divide, i.e., that there is an actual world 'out there' containing an audience and that their world is merely fictional. Fourth wall breaks can happen in variety of media (other than just theatre). Consider for instance the movie *On Her Majesty's Secret Service* which is the first (and only) movie in which George Lazenby portrayed James Bond after Bond had been portrayed by Sean Connery. In this movie, Bond is left alone on a beach holding a woman's slippers as she runs away from him. Bond says:

(5) This never happened to the other fellow

and smiles into the camera. What "other fellow" does Bond refer to? He seems to refer to Sean Connery, i.e., the actor that previously played Bond. But how can Bond possibly make reference to an actor that portrays Bond? Or consider the Deadpool comic book *Dead Presidents* which contains a panel in which the main character Deadpool states:

(6) So, I have about six pages to kill ten presidents and their henchmen. I say it's montage time. I suggest the reader crank "Five Minutes Alone" by Pantera. Perfect song for me to kick some dead president butt to.

Again, who or what is Deadpool referring to? There is no montage or reader in his world. At this point the reader may wonder whether fourth wall breaks are special to visual fictions. They are not. Although indeed a fictional character in a novel is not able to 'look the audience in the eye', they can still acknowledge their fictionality and hence break the fourth wall. For instance, consider Dickson Carr's *The Hollow Man* which is a classic example of a Locked Room Mystery novel. At some point, one of the characters, Gideon Fell, gives a lecture on the genre of Locked Room Mystery novels. When asked why this is relevant Gideon Fell answers:

(7) Because we're in a detective story, and we don't fool the reader by pretending we're not.

Before I continue, it is important to clearly distinguish the phenomenon of breaking the fourth wall (as I understand it) from a fiction in which there exists a fiction f^1 *within* the fiction f^2 and the ontological boundary between the real and fictional world is broken *within* fiction f^2. Examples of such fictions can also be found across media. For instance, in the movie *The Purple Rose of Cairo*, a woman is watching a movie when suddenly one of the fictional character walks out of the screen to interact with her. Or consider Pridelli's play *Six Characters in Search of an Author*, in which six fictional characters (abandoned by their creator) visit a theatre group to ask them to portray their story. Lastly, consider the novel *The Eyre Affair*, in which characters can enter the fictional world of literature through the use of a machine called the prose portal. Many online sources (and some theorists, e.g. Satik [2]) mention these and similar fictions as examples of (blatant) fourth wall breaks. However, I suggest that these fictions are rather simply examples of fictions in which travel between different possible worlds is possible (e.g. the *Rick and Morty* series of the movie *Everything everywhere all at once*). These may be interesting examples of impossible fictions in which it is true about one and the same entity that it *is* and *is not* a fictional entity. However, these are not examples of fictions in which a fictional character breaks the ontological barrier between the fictional world and *our* actual world.

2 Semantic analysis of BtFW-statements

Now that we have a clearer understanding of the phenomenon of breaking the fourth wall, let's consider how it relates to fictional, parafictional and metafictional discourse. I suggest that we should understand BtFW-statements as a mirror case of parafictional statements. For instance, in Bond's statement (5) "the other fellow" is taken to abbreviate something like "the other actor that portrayed the fictional character Bond". *Prima facie*, Bond is interpreted as saying something like the following:

(8) The other actor that portrayed the fictional character Bond never portrayed a Bond that women ran away from

In our world, (8) would amount to a (true) statement involving metafictional reference (i.e., reference to Bond as a fictional entity). (5) is thus a pretend assertion (made by a fictional character from a perspective internal to the fiction) that seems to involve (self-referential) metafictional (or external) reference. Or consider Dickson Carr's more explicit (7) where fictional character Gideon Fell makes metafictional reference to himself (and the people listening to his lecture) as being fictional characters by stating that they are "in a detective story".

BtFW-statements thus seem to mix internal/external perspectives in the exact opposite way that parafictional statements do. An understanding of BtFW-statements as a mirror case of parafictional statements suggests two strategies to deal with BtFW-statements that parallel the two main existing analyses of parafictional discourse, i.e., reduction to either fictional or metafictional discourse. I turn to these analyses now.

2.1 Fictional analysis

A fictional approach would analyse BtFW-statements (e.g., (5)) as a species of fictional statements. This seems consistent with our common understanding of BtFW-statements as being made by fictional characters. However, just as Bilbo does not exist in the actual world, there exists no actor that portrayed the fictional character Bond in the worlds of the Bond movies. Bond is *not* a fictional character in those worlds but a flesh and blood spy. In other words, "the other actor that portrayed the fictional character Bond" in (8) is an empty definite description in the Bond-worlds. In the fiction, (5) is thus (not even pretend-true but) a truth-valueless (or false, depending on your theory of presupposition failure) infelicitous statement made by Bond. In Bond's 'mouth', (5) is a incoherent nonsense statement. Similarly, Deadpool's "the reader of the comic book that portrays this fictional world" is an empty definite description in the *Deadpool* worlds and hence his statement (6) is similarly infelicitous.

Such an analysis seems to get part of our interpretative intuitions right. Indeed, there is a strong sense in which Bond said something nonsensical on the beach that made no sense for him to say. However, this analysis does not account for our intuition that what is expressed (i.e., that Connery never portrayed a Bond that women ran away from) is also somehow *really* true. It's not *just* nonsense.

2.2 Metafictional analysis

A metafictional approach to BtFW-statements would deem (5) a type of metafictional statement. In other words, BtFW-statements would be analysed as a special case of what Gendler [5] calls 'pop-out', i.e., an assertion by the author that 'interrupts' the fictional discourse. We are thus dealing with a temporary interruption of the pretense to make an assertion, i.e. a temporary lifting of all four walls.

In the case of (graphic) novels it is clear who takes the role of 'the author'. Similarly, in cases of pretend play or improvisational theatre, it is clear who we can ascribe individual fictional statements (and hence assertions in case of pop-out) to, i.e., the person who came up with and said the statement. Movies and scripted theatre pose an interesting intermediate case. Who is 'the author' in the case of (5)? We are dealing with an actor who (most probably) based his fictional statements on a script written by others. We can ascribe the speech act to the actor Lazenby (who was instructed to make certain fictional statements by the script writers); we can ascribe the speech act to the script writers (who make fictional statements 'through' Lazenby); or ascribe the speech act to a collective of agents including Lazenby, the writers and possibly others. For simplicitly, I am going to ascribe the relevant fictional statements (and the metafictional pop-out assertion) to Lazenby. Thus in the case of (5), Lazenby makes the (true) metafictional assertion that Connery never portrayed a Bond that women ran away from.

Again, the analysis seems to get only part of our interpretative intuitions right. Lazenby can felicitously refer to "the other actor that portrayed Bond" and hence we account for the intuition that what has been said, is really true: (5) is not nonsense in the mouth of Lazenby but a true assertion about Connery. Also, a pop-out analysis does justice to part of the aesthetic experience often associated with breaking the fourth wall, i.e., a feeling of 'interruption'. Fourth wall breaks interrupt the suspension of belief and seem to throw you out of the movie/book/fiction for a moment. However, just like the fictional analysis, the metafictional analysis also doesn't seem to do complete justice to all our interpretative intuitions. Isn't it true in the fiction that Bond says something funny/weird, i.e., that Bond says (5)? Especially in the case of the *Deadpool* comics it is clearly true in the fiction that Deadpool frequently says and does weird things. Other characters react in confusion when Deadpool breaks the fourth wall in their presence.

2.3 Definition of statements that Break the Fourth Wall

Both discussed analyses seem to get only a part of our interpretative intuitions concerning BtFW-statements right. I suggest that a synthesis of these two approaches captures the phenomenon best. I thus propose the following definition of BtFW-statements:

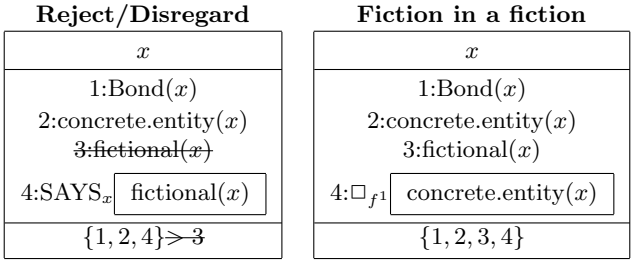
(9) Statement s in fiction f breaks the fourth wall iff s is [1] a fictional statement that makes it true in f that an infelicitous/incoherent and truth-valueless statement is made, and [2] a true and felicitous metafictional statement about f in the actual world simultaneously

For example, a BtFW-statement such as (5) instantiates two speech acts simultaneously: (5) is a fictional statement made by Lazenby that makes it true in the fiction that Bond makes a nonsensical and truth-valueless statement *and* (5) is a true metafictional assertion made by Lazenby. Or consider Dickson Carr's (7) which instantiates a fictional statement (that makes it true in *The Hollow Man* that Gideon Fell makes a incoherent statement that he and the people surrounding him are characters "in a detective story") *and* a true metafictional assertion by Dickson Carr that the relevant characters are in a detective story.

On such an analysis, there is no actual breaking of the ontological barrier between the fictional and actual world (because the rules of fiction don't allow this), but there are two parallel speech acts taking place at the same time. Apart from addressing all discussed intuitions, the proposed synthesis has the added benefit of offering an explanation for the confusing and/or funny aesthetic experience often associated with fourth wall breaks: A feeling of unexpected and awkward uncertainty on how to interpret the fiction. What's funny about BtFW-statements is that they are ambiguous between two interpretations that force a switching between internal and external perspectives. They thus put the audience in a somewhat awkward position of oscillating between two conflicting interpretations.

3 Further steps

I suggested that BtFW-statements instantiate (apart from pop-out) fictional statements that make it true in the fiction that a fictional character says something infelicitous. Put differently, the character says something that leads to an inconsistency. For instance, (8) implies that Bond is a fictional character while it is a well established fictional fact in *On Her Majesty's Secret Service* that Bond is *not* a fictional character but a flesh and blood individual. I would like to explore to what extent existing literature on interpretative strategies for inconsistency in fiction (see e.g. Matravers [9]) are applicable to fourth wall breaks. We can represent the available strategies using Maier and Semeijn's [8] ordered DRS's (Discourse Representation Structures (see Kamp [6]) that include an epistemic entrenchment ordering on the conditions):



One interpretative strategy is to reject the inconsistency by removing one of the problematic conditions. The fictional truth that is epistemically least entrenched (easiest to disregard) seems to be the 'fictional(x)' condition. This move can be accompanied by what Eckardt [3] calls a 'cautious update', i.e., I do not accept that p, but I do accept that a character believes/accepts/states that p (e.g., In *On Her Majesty's Secret Service*, Bond is not a fictional character but it *is* true that Bond states (5)). For fictions in which the fourth wall is broken frequently, it may be tempting to accept the inconsistency instead, i.e. accept that there is some entity that *can* break the ontological barrier between fiction and non-fiction and hence somehow functions on both planes of existence. Since this cannot happen in the actual world this strategy amounts to (temporarily) accommodating a fiction f^1 in the fiction. Another fourth wall is placed 'behind the audience' (e.g., the audience pretend-plays that Deadpool can actually see them). I hypothesize that excessive fourth wall breaking leads to permanent accommodation of a fiction in a fiction interpretation and hence at some point disables further fourth wall breaks.

Second, because this paper focuses on the mirror case of parafictional statements, the provided definition (9) deals only with *assertions* through which characters break the fourth wall. However, just as other kinds of parafictional speech acts are possible (e.g., "Is it true in *The Lord of the Rings*, that Bilbo is Frodo's uncle?"), characters can also break the fourth wall with other speech acts than assertions (e.g., "Would this ever have happened to the other fellow"). A general definition of BtFW-discourse could be formulated along the following lines:

(10) Discourse s in fiction f breaks the fourth wall iff s [1] is fictional discourse that makes it true in f that someone engages in infelicitous discourse, and [2] felicitous metafictional discourse about f in the actual world simultaneously

Even broader, characters can also break the fourth wall without using any language (e.g., in the videogame *Sonic the Hedgehog*, the avatar will look at the player impatiently in case they don't press any buttons for too long). In order to do justice to all varieties of fourth wall breaks in terms of (meta)fictional reference, we would need an account of reference that can for instance analyse a meaningful look into a camera as a kind of pointing or demonstrative reference.

Last, it would be interesting to see how the provided analysis relates to other interesting phenomena such as 'leaning against the fourth wall', 'fourth wall psychs' and 'asides'.

References

[1] Diderot Denis. On dramatic poetry. In B. H. Clark, editor, *European Theories of the Drama*. Crown Publishers, New York, 1947 [1758].

[2] Satik Deniz. The fourth wall against possibilism on truth-in-fiction (manuscript). 2021.

[3] Regine Eckardt. *The Semantics of Free Indirect Discourse*. Brill Publishers, Leiden, 2014.

[4] Gareth Evans. *The Varieties of Reference*, volume 10. Oxford University Press, Oxford, 1982.

[5] Tamar Gendler. Imaginative resistance revisited. In Shaun Nichols, editor, *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction*, pages 149–173. Oxford University Press, Oxford, 2006.

[6] Hans Kamp. A theory of truth and semantic representation. In Jeroen A. G. Groenendijk, Theo M. V. Janssen, and Martin B. J. Stokhof, editors, *Formal Methods in the Study of Language, Part 1*, pages 277–322. Blackwell Publishers Ltd, Oxford, 1981.

[7] Tobias Klauk. Zalta on encoding fictional properties. *Journal of Literary Theory*, 8(2):234–256, 2014.

[8] Emar Maier and Merel Semeijn. Extracting fictional truth from unreliable sources. In Emar Maier and Andreas Stokke, editors, *The Language of Fiction*, pages 186–220. Oxford University Press, Oxford, 2021.

[9] Derek Matravers. *Fiction and Narrative*. Oxford University Press, Oxford, 2014.

[10] Semeijn Merel. The 'in' and 'according to' operators. In *Proceedings of the ESSLLI & WeSSLLI Student Session 2020*. 2020.

[11] François Recanati. Fictional, metafictional, parafictional. *Proceedings of the Aristotelian Society*, 118(1):25–54, 2018.

[12] Merel Semeijn and Edward N. Zalta. Revisiting the 'wrong kind of object' problem. *Organon F*, 28(1):168–197, 2021.

[13] Nele Van de Mosselaer. Breaking the fourth wall in videogames. In Enrico Terrone and V. Tripodi, editors, *Being and Value in Technology*. Palgrave Macmillan, Cham, 2022.

[14] Peter van Inwagen. Quantification and fictional discourse. In Anthony Everett and Thomas Hofweber, editors, *Empty Names, Fiction and the Puzzles of Non-Existence*, pages 235–47. CSLI, Stanford, 2000.

[15] Edward N. Zalta. *Abstract Objects: An Introduction to Axiomatic Metaphysics*. Springer, Berlin, 1983.

Projective variability of (semi)factive verbs in family of sentence contexts: A rating study*

Judith Sieker and Torgrim Solstad

Bielefeld University
{j.sieker, torgrim.solstad}@uni-bielefeld.de

Abstract

Investigating the influence of the embedding context on the projectivity of factive and semifactive predicates, this study has two main goals. On the one hand, we investigate empirically whether factive and semifactive verbs exhibit different degrees of projectivity depending on the Family of Sentences context in which they are embedded (as claimed by Karttunen [9]). Furthermore, we also investigate the relationship between projectivity and at-issueness for these contexts, thereby offering new experimental evidence regarding the Gradient Projection Principle (as formulated by Tonhauser et al. 2018 [23]). The results from our web-based rating study in German show two things: First, our study provides new empirical evidence that the Gradient Projection Principle does indeed generalize to the other Family of Sentences contexts. Second, our results do not support [9]'s theoretical claim as to the influence of the embedding context on projectivity for these verbs. Still, the data provide new empirical support for the observation that the individual verbs do differ with regard to their projectivity, beyond their partition along the lines of (semi)factivity.

1 Projective Variability among (semi)factive verbs

Among the semantic dimensions along which presuppositions may be distinguished, projection, that is, the ability to survive in an entailment-cancelling environment, is undoubtedly among their most prominent characteristics. However, despite being used as a means to identify presuppositions (e.g. [4], [2]), existing theoretical and experimental research has shown that presuppositions do not project equally stably in these environments (e.g. [23], [7], [26]). This projective variability can be attributed to various factors, such as prosody and information structure (e.g. [22], [8], [19]), personal beliefs (e.g. [13], [6]) or the particular environment in which a presupposition is embedded. The impact of the latter on the projective behaviour of factive verbs was already observed by Karttunen [9], as discussed for his original examples (1):

(1) a. John didn't {regret / realize / discover} that he had not told the truth.
b. Did you {regret / realize / discover} that you had not told the truth?
c. If I {regret / realize / discover} later that I have not told the truth, I will confess it to everyone.
d. It is possible that I will {regret / realize / discover} later that I have not told the truth.

According to [9], while each verb is projective when embedded under negation (1a), that is, each verb triggers the presupposition that the agent of the complement has not told the truth, this is no longer the case in the other environments in (1b) - (1d). In particular, *realize* and *discover* are claimed to differ from *regret* as they allow for both factive and non-factive-interpretations. The

*We wish to thank Oliver Bott for the very helpful discussions, suggestions and support throughout this study.

presupposition of the former can, but must not, project when embedded in polar questions, in the antecedents of conditionals, or in modal constructions. This observed projective variability was Karttunen's impetus to differentiate between factive and semifactive verbs. Thus, [9], as many others following him (e.g. [1], [7]), posited the thesis that factive verbs do not form a homogeneous class. What is notable about his account, however, is that the projective variability among factive verbs is taken to be a function of Family of Sentence ([4]; henceforth FoS) context, which distinguishes his approach from other categorizations of factive verbs. Existing empirical work ([23], [7] or [26]) was able to identify projective variability between factive verbs, but did not examine to what extent the FoS environment has an impact. The present study therefore investigates whether the variability in projection between factives and semifactives depending on the FoS environment can be empirically confirmed, as theoretically hypothesized by [9].

To examine the projective behaviour of factive and semifactive verbs relative to FoS context, we employed the experimental paradigm used by [23], who studied the projectivity and at-issueness of various presupposition triggers in polar questions, and extended it to the other three FoS contexts. This naturally leads to another objective for our study: The investigation of the relationship between projectivity and at-issueness, that is, the extent to which content is considered to be up for debate (e.g. [15], [21]). For instance, [23], among others (e.g. [26], [12]), were able to provide strong empirical evidence that at-issueness is a promising explanation for the variability in projection between presupposition triggers. This was formulated by [23] in their *Gradient Projection Principle*, stating that a presupposition projects more strongly the less at-issue it is. However, in which way the Gradient Projection Principle may also interact with the FoS context has, to our knowledge, not been considered. Our study therefore set out to replicate earlier findings on the close relationship between the degrees of (non-)at-issueness and projectivity in polar questions and, furthermore, to explore to what extent these can be transferred to the other three FoS contexts as discussed for the examples in (1).

2 Empirical Investigation

2.1 Methods

In this experiment ($N = 105$; participants recruited via Prolific), the potential projective variability as hypothesized by [9] was investigated for six different factive or semifactive verbs in German. The three factive verbs *wissen* 'know', *bereuen* 'regret' and *enthüllen* 'reveal' were chosen as they have been found to show particularly strong projective behaviour in other (theoretical and empirical) studies (e.g [11], [23], [7]). The three semifactive verbs *bemerken* 'notice', *entdecken* 'discover' and *herausfinden* 'find out'/ 'learn' were chosen primarily for the reason that they – or, rather, their English equivalents – were explicitly classified as semifactive by [9].

To elicit **projectivity**, we employed the *certain that* test of [23], which measures projectivity by assessing speaker-commitment and which has proven to be a robust method, e.g. by [22], [8], [20], [13] or [7]: Participants were asked to imagine overhearing utterances as part of brief dialogues at a kitchen party. In the utterances the expressions associated with projective content were embedded in an FoS environment such as a question: "Did Marius reveal that Kim defrauded the tax office?". Participants' task was to indicate by means of a slider from "no" (0) to "yes" (100) whether they would judge the speaker to be certain about this particular content: "Is the speaker certain that Kim defrauded the tax office?". Projection should evoke a positive answer.

Based on [9]'s observations for (1), we extended the design in [23] by including three more FoS

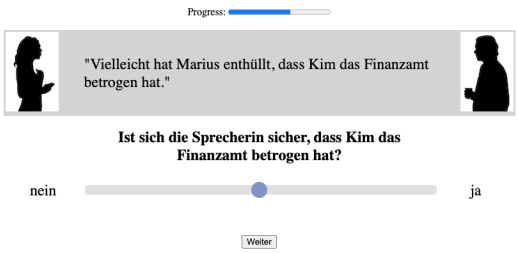


Figure 1: Sample experimental trial of the *certain that* test for the trigger *enthüllen* 'reveal' in a modal context from the internet questionnaire with the respective translation given in (2).

contexts in addition to questions: embedding under negation, in the antecedents of conditionals, and in modal constructions with *vielleicht* 'maybe'. Accordingly four different experimental conditions were realized in the *certain that* test, as expemplified for the factive *enthüllen* 'reveal':

(2) **Projectivity Ratings**

a. **question:** The speaker asks: "Did Marius reveal that Kim defrauded the tax office?"

b. **negation:** The speaker says: "Marius did not reveal that Kim defrauded the tax office."

c. **antecedence of conditional:** The speaker says: "If Marius revealed that Kim defrauded the tax office, then he will not trust her with any documents."

d. **modal:** The speaker says: "Maybe Marius revealed that Kim defrauded the tax office."

Question: Is the speaker certain that Kim defrauded the tax office?

An example trial for the modal FoS context is shown in Figure 1.

In order to elicit the **at-issueness** of the same verbs, we adopted the *Are you sure?* method from [23]'s Experiment 2, measuring the extent to which a presupposition is up for debate by asking the participants to judge (again, using a slider ranging from "no" to "yes") whether the presupposition addresses a preceding "Are you sure?" question. Here, at-issueness should evoke a negative answer. Departing from the design of [23], the at-issueness of the triggers was only investigated in simple assertions, that is, in one experimental condition, as exemplified in (3) for the semifactive verb *herausfinden* 'find out'/'learn', as we did not see how to apply the *Are you sure?* method in the scope of entailment-canceling operators, that is, in FoS contexts.

(3) **At-Issueness Rating**

a. The speaker says: "Pia found out that David has British ancestors."

b. Her interlocutor asks: "Are you sure?"

c. The speaker says: "Yes, I am sure that David has British ancestors."

Question: Did the speaker answer her interlocutor's question?

For each of the six factive and semifactive verbs, 15 different lexical contexts/situations were constructed, each of which occurred in the five different conditions of the experiment (as shown in (2) and (3)). The resulting items were distributed over five experimental lists according to a Latin Square design (90 trials in each list). In addition to the factive and semifactive verbs, our experiment included another presupposition trigger type, the so-called *Occasion verbs* (cf. [3, 18]) as filler items (15 items, each in the five conditions). Additionally, each list included 30

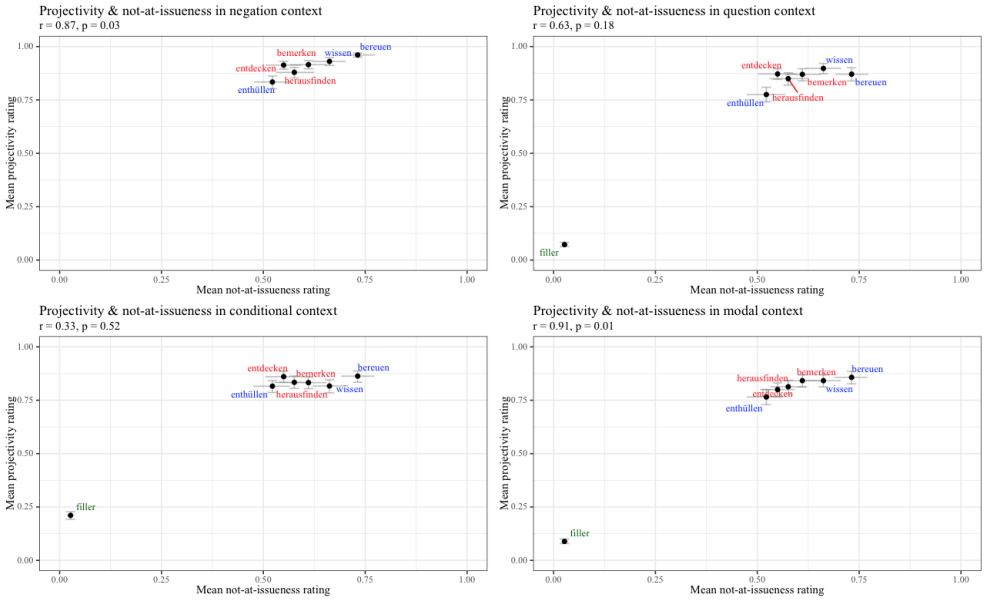


Figure 2: Projectivity/(non)-at-issueness ratings per FoS context and correlation coefficients between not-at-issueness and projectivity ratings (95% confidence intervals estimated via bootstrapping); factive verbs in blue, semifactives in red, distractor items in green.

distractor items querying about clearly non-projective content (in all FoS contexts except for negation) as well as at-issueness to check whether participants were focusing on the task.

2.2 Results and Discussion

We will first discuss whether the Gradient Projection Principle generalizes to the other FoS contexts. The plots in Figure 2 show the mean projectivity and not-at-issueness ratings of the individual triggers in each of the four FoS environments. The observed positive correlations between non-at-issueness and projectivity suggest an effect of at-issueness for each of the FoS contexts: higher non-at-issueness ratings resulted in higher projectivity ratings, as predicted by the Gradient Projection Principle. Put differently, there seems to be a clear relationship between (non)-at-issueness and projectivity across FoS contexts. This observation was corroborated by linear mixed-effects analyses, finding a significant main effect of at-issueness on projectivity in each of the FoS contexts (question: $\chi^2(1) = 4.09, p < 0.05$; negation: $\chi^2(1) = 5.56, p < 0.02$; antecedent of conditionals: $\chi^2(1) = 11.67, p < 0.001$; modal construction: $\chi^2(1) = 7.26, p < 0.01$). Further, when aggregated over the four FoS contexts, linear mixed-effects analyses found a significant main effect of at-issueness on projectivity ($\chi^2(1) = 9.04, p = 0.003$), but no significant interactions between at-issueness and FoS context ($\chi^2(3) = 7.17, p = 0.07$). Hence, these results not only support [23]'s findings on the Gradient Projection Principle in polar questions, but also provide new empirical evidence that this principle does generalize to the other FoS environments.

Next, we will consider the empirical applicability of [9]'s observation regarding the differences in projective variability between factive and semifactive verbs. As can be seen from Figure 3, the data did not show projective variability between the two verb classes depending on the

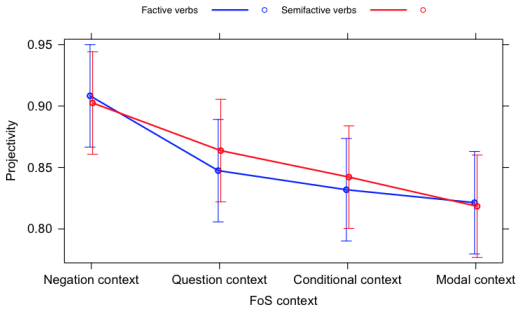


Figure 3: Relationship between FoS context and trigger type on projectivity.

FoS environment. Rather, the two trigger types showed a similar projective behaviour in each of the FoS contexts (recognizable by their lines running parallel across the contexts). What is more, if we take another look at Figure 2, it can be seen that the individual verbs cluster very similarly in all FoS contexts. This observed lack of variability across contexts was confirmed by a linear mixed-effects regression model analyzing the fixed effects of at-issueness and FoS environment, revealing no significant interaction effects ($\chi^2(3) = 3.27, p = 0.35$). Moreover, across FoS contexts, both factive and semifactive verbs were rated to be rather highly projective and, more importantly, did not differ noticeably in their projectivity ($\bar{x}_{factives} = 0.85$, $\bar{x}_{semifactives} = 0.86$), hence turning out to be more uniform than assumed. This was also supported by linear mixed-effects analyses, revealing no significant main effect of trigger type on projectivity ($\chi^2(1) = 0.03, p = 0.85$).

The findings of our study can thus be taken to speak against a substantially different modulation of projective variability depending on the FoS context for factive and semifactive verbs. However, we did find projective variability between FoS contexts, as can be seen, for example, by the descending lines in Figure 3. This observation found support by a linear mixed-effects regression model, revealing a significant main effect of FoS context on projectivity ($\chi^2(3) = 158.45, p < 0.0001$). Furthermore, our findings provide evidence for variability in the projective behaviour of the individual verbs, as follow-up analyses found a significant main effect of the individual verbs on projectivity ($\chi^2(5) = 50.56, p < 0.0001$). In particular, the verb *enthüllen* 'reveal' differed significantly from the other factive verbs in that it was always the least projective trigger across all the FoS environments, that is, even less projective than the semifactive verbs (as can also be seen in Figure 2).

3 General Discussion

The main findings of our study can be summarized as follows. First, [9]'s theoretical observation regarding the projective variability between factive and semifactive verbs depending on FoS context could not be confirmed empirically. Second, we found projective variability across FoS contexts as well as across individual verbs, when detached from their classification along the lines of (semi)factivity. Finally, and most importantly, our study is – to our knowledge – the first to show that [23]'s Gradient Projection Principle was found to be transferable to the other FoS contexts.

Our results support the findings in previous research. With regard to projective variability for different FoS contexts, for instance, [16, 17] also found conditionals and negation to differ

with regard to the strength of their projectivity. It should, however, be added that their projectivity ratings do not entirely coincide with ours. One desideratum for future research is therefore to consider possible reasons for the different projectivity ratings across FoS contexts. Furthermore, while the (semi)factive verbs in the MegaVeridicality dataset ([24, 25]) would indeed be in line with [9]'s observation, our conclusion that [9]'s differentiation into factive and semifactive verbs cannot be empirically proven, finds support in the results of [26], [23] and [7] for the (semi)factive verbs included in their investigations as well as for the (semi)factive verbs in the VerbVeridicality ([14]) and CommitmentBank datasets ([5]). However, these studies found, as we did, projective variability between individual verbs. Of particular interest is the fact that [23] and [7] also found the verb corresponding most closely to *enthüllen*, *reveal*, to exhibit a weaker projective behaviour in comparison to the semifactive verbs that were studied there, even though the individual verbs were not categorically distinguishable from each other in terms of their projectivity. This supports observations that there are other factors besides FoS context that may also have an influence on projectivity (see, for instance, [5] for discussion). The following considerations seem to be of relevance. For instance, the findings of [13] or [6] have shown that the prior beliefs of listeners can have an impact on the extent to which a presupposition projects. In these studies, contents that were presented with facts higher in probability also received higher projectivity ratings. Consequently, it may be that the party context in which all the items were embedded (cf. Section 2.1) was judged to be less plausible for the verb *enthüllen* compared to the other verbs, since the revelation of a fact implies that this very fact was not (publicly) known beforehand. This in turn might lead to higher uncertainty ratings and thus lower projectivity. This assumption is reminiscent of another factor that may also contribute to the salience of the verb *enthüllen*, which is the notion of factive verbs as (un)informative presuppositions (e.g. [19]). In this context, the complement of the verb *enthüllen* could be comprehended as higher in informativity than, for instance, the one for *know*, thereby being more at-issue and hence less projective than the other, not (so) informative presuppositions.

Finally, it may be considered whether the differences in projectivity could be due to differences in argument structure, a factor not commonly associated with projectivity. Thus, in our experiment, *enthüllen* 'reveal' was the only three-place verb including an argument position for the addressee/recipient, as in "x reveals to y that z". As our items were always of the type "x reveals/knows/regrets that z" (see (2)), *reveal* would always occur without the addressee of the revelation being specified. In this case, however, it does not seem too far-fetched that the speaker in our experimental items could be implicitly included among the addressees of the revelation. For instance, one might imagine "Marius revealed that Kim defrauded the tax office", being interpreted as "Marius revealed to us that Kim defrauded the tax office". If this were to be the case, however, it may well that filtering is enforced, as there would otherwise be a risk of incoherence, compare *#Kim defrauded the tax office, and Marius revealed to us that Kim defrauded the tax office.* Should these speculation lead in the right direction, we would expect different explicitly mentioned addresses to be associated with projective variability across all FoS contexts as illustrated in (4):

(4) a. Marius revealed to us that Kim defrauded the tax office.
b. Marius revealed to them that Kim defrauded the tax office.

There are certainly a number of additional factors or issues which may influence the projective variability of factive and semifactive verbs. In any case, joining the effort to clean up the mess so "joyfully" left behind by Karttunen and colleagues ([10]: p. 728) proves time and again to be an intriguing and entertaining endeavour.

References

[1] David Beaver. Have you noticed that your belly button lint colour is related to the colour of your clothing? In Rainer Bäuerle, Uwe Reyle, and Ede Zimmermann, editors, *Presuppositions and Discourse: Essays Offered to Hans Kamp*. Elsevier, Oxford, 2010.

[2] David Beaver, Bart Geurts, and Kristie Denlinger. Presupposition. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition, 2021.

[3] Oliver Bott and Torgrim Solstad. Cataphoric presuppositions verification and symmetric filtering: the case of occasion verbs. In *Proceedings of Amsterdam Colloquium 2022*, Amsterdam, 2022. ILLC.

[4] Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar. An Introduction to Semantics*. MIT Press, Cambridge, 1990.

[5] Marie-Catherine de Marneffe, Mandy Simons, and Judith Tonhauser. The CommitmentBank: Investigating projection in naturally occurring discourse. In *Sinn und Bedeutung*, volume 24, pages 107–124, 2019.

[6] Judith Degen and Judith Tonhauser. Prior Beliefs Modulate Projection. *Open Mind*, 5:59–70, 2021.

[7] Judith Degen and Judith Tonhauser. Are there factive predicates? An empirical investigation. *Language*, 3(98), 2022.

[8] Kajsa Djärv and Hezekiah Bacovcin. Prosodic Effects on Factive Presupposition Projection. *Semantics and Linguistic Theory*, 27:116–133, 2017.

[9] Lauri Karttunen. Some observations on factivity. *Paper in Linguistics*, 4(1):55–69, 1971.

[10] Lauri Karttunen. Presupposition: What went wrong? *Semantics and Linguistic Theory*, 26:705–731, 2016.

[11] Paul Kiparsky and Carol Kiparsky. FACT. In Manfred Bierwisch and Karl Ehrich Heidolph, editors, *Progress in Linguistics: A Collection of Papers.*, pages 143–173. Mouton, 1971.

[12] Alexandra Lorson. The influence of world knowledge on projectivity: Listeners revise their gender stereotypes when processing presuppositions triggered by stop. Master's thesis, The University of Potsdam, 2018.

[13] Taylor Mahler. The social component of projection behavior of clausal complements. *Linguistic Society of America*, 5(1):777–791, 2020.

[14] Alexis Ross and Ellie Pavlick. How well do NLI models capture verb veridicality? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 2230–2240. Association for Computational Linguistic, 2019.

[15] Mandy Simons, Judith Tonhauser, David Beaver, and Craige Roberts. What projects and why. In *Proceedings of SALT*, volume 20, pages 309–327, 2010.

[16] Allyn E. Smith and Kathleen Currie Hall. Projection Diversity: Experimental Evidence. In *Proceedings of the 2011 ESSLLI Workshop on Projective Content.*, pages 156–170, Columbus, OH, 2011.

[17] Allyn E. Smith and Kathleen Currie Hall. The relationship between projection and embedding environment. In *Proceedings of the 48th Meeting of the Chicago Linguistics Society*, Chicago, 2013. University of Chicago.

[18] Torgrim Solstad and Oliver Bott. Cataphoric presuppositions: The case of occasion verbs. Unpublished manuscript., 2022.

[19] Jennifer Spenader. Factive Presuppositions, Accommodation and Information Structure. *Journal of Logic, Language, and Information*, 12(3):351–368, 2003.

[20] Jon Scott Stevens, Marie-Catherine de Marneffe, Sharie R. Speer, and Judith Tonhauser. Rational

use of prosody predicts projection in manner adverb utterances. *Cognitive Science*, pages 1144–1149, 2017.

[21] Judith Tonhauser. Diagnosing (not-)at-issue content. *Semantics of Under-represented Languages of the Americas*, 6:239–254, 2012.

[22] Judith Tonhauser. Prosodic cues to presupposition projection. *Semantics and Linguistic Theory*, 26:934–960, 2016.

[23] Judith Tonhauser, David Beaver, and Judith Degen. How Projective is Projective Content? Gradience in Projectivity and At-issueness. *Journal of Semantics*, 35(3):495–542, 2018.

[24] Aaron S. White and Kyle Rawlins. The role of veridicality and factivity in clause selection. In *48th Meeting of the North East Linguistic Society*, pages 221–234, 2018.

[25] Aaron S. White, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. Lexicosyntactic inference in neural models. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4717–4724, Brussels, 2018. Association for Computational Linguistics.

[26] Jingyang Xue and Edgar Onea. Correlation between Presupposition Projection and At-issueness: An Empirical Study. In *Proceedings of the 2011 ESSLLI Workshop on Projective Content.*, pages 171–184, Columbus, OH, 2011. The Ohio State University.

Bare nominals are underspecified for (in)definite meanings, not ambiguous between them*

Sadhwi Srinivas[1]

College of William & Mary, Williamsburg, Virginia, USA
ssrinivas@wm.edu

Abstract

Definite bare nominals in *article-less* languages exhibit a more limited distribution than English *the* in anaphoric contexts. Here, I argue that a comprehensive account of this fact is facilitated by a wherein bare nominals are *underspecified* for (in)definiteness in episodic contexts, instead of the standard view that they are *ambiguous* – with the two meanings arising through completely independent means. I discuss how assuming underspecificity in the bare nominals opens up a new type of empirically viable, analytically parsimonious account of their anaphoric variability unavailable under the ambiguity view.

1 Introduction

Delineating the exact relationship between the *form* of a nominal expression and the range of meanings it is compatible with is a non-trivial task even within a single language, and markedly more complicated when we look across languages. Imagining the space of nominal meanings as a continuum, this task may be construed as one that demarcates possibly overlapping regions of the continuum expressible by one or other nominal form within and across languages. In crosslinguistic studies, it is often assumed that identical meaning demarcations are instantiated across languages, even if they are not mapped to overtly distinct lexical forms in all of them. For example, some recent accounts take both German and English to express (at least) two types of definite meanings: a uniqueness-based meaning and an anaphoric one, though German alone overtly distinguishes between the forms expressing these two meanings [22] while English employs *the* for both [1, 19].

So-called *article-less* languages that lack overt counterparts of articles corresponding to English *the* and *a* pose a unique challenge to the crosslinguistic form-to-meaning mapping task. On the one hand, it appears that article-less 'bare' nominals in these languages span across demarcations in the 'nominal meaning continuum' alluded to above, in that they are capable of expressing both existential and definite-like meanings (e.g., [6, 9, 10, 11, 16]), translatable using English *a* and *the* respectively.[1] On the other hand, both their existential and definite-like uses have been crosslinguistically noted to be more limited when compared to English counterparts *a* and *the* respectively (e.g., [18, 19, 1, 2, 4, 3] *i.a* all note the limited definite interpretations of bare nominals in various languages when compared to English *the*; [9, 10, 13, 7] *i.a* note that the existential interpretations of bare nominals are restricted to narrow scope unlike English *a*). It is thus a challenge to precisely map out and explain the range of occurrences of the bare nominal in its various senses, especially in comparison to its crosslinguistic counterparts.

*I am grateful to Kyle Rawlins, the audiences at the *New York Philosophy of Language Workshop*, three anonymous reviewers at AC 2022, and the linguistics research group at William & Mary for discussion and feedback. All errors are my own.

[1]Bare nominals are also generally compatible with kind readings in *article-less* languages, but here we will focus our attention only to episodic contexts that do not license kind interpretations.

Historically, the study of the definite-like uses of bare nominals has tended to be carried out separately from its existential uses, with studies focusing either on what makes the bare nominal unable to interpreted as a definite description in certain contexts where *the* is licensed, or on the differences between the indefiniteness potential of the bare nominal compared to *a*, but not both at once.[2] In the current paper, I too will focus on the distribution of bare nominals in their definite-like uses, though the conclusion I will defend is that this investigation cannot be fruitfully carried out while completely neglecting their existential uses. In other words, while confining ourselves to examining where the definite article *the* (but not *a*) is licensed makes sense in the study of definiteness marking in English, it is less clear that this is a cogent move in *article-less* languages – which do not themselves make this distinction. Instead, I will argue, it is more constructive to move away from assumptions that stem from an English-like picture in investigating bare nominals in *article-less* languages.

The question of the distribution of definite-like uses of bare nominals has generated much discussion in recent semantic work. In particular, it has been noted that definite interpretations of bare nominals are limited in anaphoric contexts when compared to English *the*, as seen in the Mandarin example (1) below taken from [19]:

(1) Jiaoshi li zuo zhe yi ge nansheng yi ge nüsheng. Wo zuotian yudao
classroom inside sit PROG one CL boy one CL girl I yesterday meet
#(**na** *ge*) nansheng.
(that CL) boy
"There is a boy and a girl sitting in the classroom. I met the boy yesterday."
([19]: ex. 15a-b)

Though several proposals have been advanced to explain what restricts the definite interpretations of bare nominals in cases like (1), none of these quite paints the full picture, as I discuss in Section 2. In Section 3, I locate a common (often implicit) assumption made by all these accounts as the source of their shortcomings: namely, one of ambiguity in the bare nominal in whether it expresses an existential or definite-like meaning. I discuss some of the (arguably non-ideal) methodological and analytical impacts of this assumption on the study of (in)definiteness in the bare nominal. Following this, in Section 4, I advocate an alternative analysis of the bare nominal in which it is *underspecified* for (in)definiteness. Section 5 concludes.

2 Existing Analyses

The limited definite-like uses of bare nominals in anaphoric contexts as in (1) has been noticed not only for Mandarin, but also for several other *article-less* languages: Thai [18], Hindi [11], Akan [3], Korean [1, 2], Kannada [24, 23] *i.a.* In this section, I discuss two types of explanations proposed to account for such anaphoric variability, and discuss some problems they run into.

Jenks 2018. Inspired by the ideas in Schwarz (2009) that languages can separately instantiate (and potentially overtly distinguish between) uniqueness-based and anaphoric definiteness, [19] advances an analysis of definiteness in Mandarin whereby the definite meanings arising with article-less 'bare' nominals are uniqueness-based, while the demonstrative form in this language instantiates anaphoricity. Such a proposal finds tangential support in the observation that definite readings of Mandarin bare nouns are limited in anaphoric contexts in comparison to *the* (taken to be ambiguous between the uniqueness and anaphoric meanings), but it does not

[2] Dayal's work on Hindi bare nouns (Dayal 1992, 1999, 2004, 2011) is arguably an exception to this generalization. Her work provides one of the first detailed descriptions of bare nominals in an *article-less* language. As such, she discusses data corresponding to the full range of their uses, though her analysis is focused on the relationship between the kind and existential uses of the bare nominals and less so on their definite-like uses.

directly account for why the bare nominal is not independently licensed by the existence of a unique boy in the discourse context of (1), regardless of anaphoricity.

To altogether rule out the definite reading of the bare noun in (1),[19] proposes a specialized constraint *Index!*, based on *Maximize Presupposition!* [15], by which the anaphoric demonstrative form associated with a stronger presupposition *must* be used when licensed, to the detriment of definite bare nominals in these contexts (barring a stipulated exception when the bare noun is the sentence-initial 'Topic').[3] However, *Index!* proves too strong to capture the actual distribution of bare nouns in Mandarin – as discussed in recent responses by [12] and [4]. Both responses note that Mandarin bare nouns in non-sentence-initial positions do receive definite interpretations in many naturally occurring anaphoric contexts; see for instance (2):

(2) Jiaoshi li zuo zhe yi ge nansheng yi ge nüsheng. **nüsheng** zuo zai
classroom inside sit PROG one CL boy one CL girl girl sit DUR
nansheng pangbian.
boy side
"There is a boy and a girl sitting in the classroom. The girl was sitting next to the boy."
([12]: ex. 12a-b)

Bremmers et al. 2021. Following a corpus examination of the contexts in Mandarin where bare nominals can and cannot receive definite interpretations, [4] argue for a hypothesis that distinguishes two types of anaphoric (or familiarity-based) definiteness meanings across languages: the standard notion of *text-level familiarity* where the antecedent has been previously introduced as part of the discourse, and a stricter notion of *situation-level familiarity* where the antecedent has been previously introduced in the same *situation* as the subsequent mention. In particular, they propose that Mandarin bare nouns receive definite interpretations in contexts that instantiate *situation-level familiarity* – but *text-level familiarity* does not suffice on its own. Overall, their view of Mandarin bare nouns is that they are "ambiguous between weak and strong definites, with the latter restricted to situation-level familiarity."

The view espoused in [4] accounts for the Mandarin data points (1) and (2) in the following way. In (1), the bare nominal *nansheng* ('boy') in the second sentence is situated *yesterday*, but the antecedent boy was introduced in a separate *classroom* situation. In this case, while text-level familiarity holds, the shift in the situations that the two sentences are *about* results in a violation of situation-level familiarity. As such, the bare nominal does not receive a definite interpretation. By contrast, in (2), the subsequent sentence containing the bare nominal continues to describe the *classroom* situation in which the antecedent was introduced, thus satisfying situation-level familiarity and thereby permitting a definite interpretation.

In its ability to anticipate and concretely explain the contrast between these two examples in Mandarin, this account is superior to its predecessors. Moreover, the core idea behind this proposal – that the anaphoric interpretation of the bare nominal is sensitive to the identity of the sentential situation – is likely to be on the right track.[4] However, we take issue with the specific implementation of this idea by these authors, whereby a new crosslinguistic sub-type of definiteness is introduced. For one thing, the standard universality assumption, according to which identical meaning distinctions are underlyingly present in every language even if not

[3] [2] presents a different analysis of the anaphoric variability of bare nominals, but resembles [19] in positing a categorical economy constraint in which a more economical form competes with the bare nominal in anaphoric contexts (in that case, the simplex pronominal form) and completely rules it out in some languages.

[4] [12] also favor a situation-sensitive semantics for the Mandarin bare noun, though they refrain from fully developing it due to their view that "further theorizing has to wait till we have more data to formulate sound empirical generalizations on". The current work does not disagree with the spirit of their conclusion, but only presents an alternative framework for hypothesis formation in future empirical work.

overtly distinguished, predicts that situation-level familiarity *must* be instantiated in every language. However, in a language like English where the weaker notion of text-level familiarity suffices to license *the*, claiming that *the* further corresponds to situation-level familiarity is entirely redundant as contexts that instantiate situation-level familiarity also entail text-level familiarity. Alternatively, to say that English does not instantiate situation-level familiarity violates the universality assumption, potentially permitting a proliferation of crosslinguistic definite meanings only subsets of which are realized in each language.[5]

Furthermore, if the Mandarin bare noun is ambiguous between a uniqueness and a situational familiarity meaning, it is unclear why a uniqueness-based definite interpretation of the bare nominal in (1) cannot independently arise, since uniqueness in the discourse context is ostensibly satisfied, even if situation-level familiarity is not. As discussed in [22] (ex. 49-50), such contexts can license the uniqueness-denoting form of the definite article in German. Finally, the categorical nature of situation-level familiarity as discussed by [4] predicts the definite reading of the bare nominal to be unavailable whenever there is a shift in the antecedent *vs.* subsequent situations. However, in some *article-less* languages like Kannada, the definite reading of the bare nominal is significantly (though not categorically) improved despite a shift in the narrative situation if there is enough salient context surrounding the antecedent in prior discourse:

(3) Nenne aDigemane-alli ondu ili ooDaaDutta ittu. Adu nooDakke tumbaa
Yesterday kitchen-in one mouse roaming was it to.see very
muddaag-ittu. Adara moogina hattra chikka macche ittu. Ivattu
cute-be.PST its nose-GEN near small birthmark be.PAST today
baccalamane-alli kooDa ?**ili** kaaNisitu.
bathroom-in too mouse seen.CAUS
'Yesterday, in the kitchen, I saw a mouse$_k$. It was really cute. There was a small birthmark near its nose. Today in the bathroom too, the mouse$_k$ was roaming around.'

3 A Deep Assumption of Ambiguity

The goal of this section is to elucidate a common working assumption underlying existing approaches to explaining the distribution of definite-like bare nominals: that the existential and definite-like meanings arise through independent considerations. In other words, the bare nominal is assumed to be practically ambiguous between its (in)definite meanings. This assumption is often not explicitly stated, and perhaps not even intended by the authors in its crudest sense where there are distinct entries in the mental lexicon corresponding to the existential and definite-like bare nominals. Nevertheless, two key reflections about the methodology adopted and the nature of analyses proposed in these works give its lurking presence away.

First, studies on definiteness in the bare nominal, including the ones considered in Section 2, restrict themselves to contexts that license the definite article in languages like English but not the indefinite one, despite no evidence internal to the *article-less* language under study itself that such contexts should be separated analytically – the reverse, if anything, seems to be the case with respect to these languages. I suggest that such a solely comparative approach to the definite-like uses of bare nominals limits how far we can get in explaining their full distribution. It is perhaps helpful as a point of comparison to consider that the study of English *the* has suffered no such disadvantage, and plentiful (attempts at) unified alternatives to the semantics of the definite article that span across their various uses[6] exist alongside the

[5] While there is no inherent inconsistency in a world where a strong version of universality is false, it diverges from what is standardly assumed in the literature, and needs thinking through all its consequences.

[6] with the exception of *weak definites*; see for example, [5]

recent ambiguity view that there are two '*the*'s, one instantiating uniqueness and the other familiarity. By contrast, no serious unified hypothesis for the bare nominal exists that spans across its existential and definite-like uses in episodic contexts.[7]

On the analytical side, treating the existential and definite-like uses of bare nominals as entirely separate – a treatment warranted under an ambiguity-based view – necessitates positing specialized constraints to independently rule out definite interpretations in anaphoric contexts before the indefinite meaning can arise. In other words, the ambiguity assumption together with a constraint like *Maximize Presupposition!*, precludes an explanation in which the *stronger*, definite sense of the bare nominal directly competes with the presupposition-less indefinite sense and loses (cf. [8]). Such categorical and independent constraints are exactly what we find in existing proposals. including *Index!* in [19] and the notion of situation-level familiarity in [4].

4 An Underspecified Bare Nominal

If we abandon the independence assumption and instead adopt the view that bare nominals are *underspecified* for (in)definiteness[8], an account where the two senses compete more equally is now open to us, where the indefinite sense may be naturally preferred in certain contexts. In this proposal, which of the two meanings is obtained depends on the outcome of a probabilistic, context-sensitive process I sketch below, which only uses an independently motivated view of domain restriction in addition to the underspecification assumption. Crucially, an underspecified bare noun does not lexically specify any presuppositions, making *Maximize Presupposition!* inapplicable, so that the indefinite and definite meanings can battle it out on a par.

Assuming, as is standard, that the bare nominal introduces an abstract discourse referent (or index), I follow [14] in mapping indices to entities drawn from potentially different types of *value sets*. In this framework, underspecification can be implemented as uncertainty in the value set to which the index introduced by the bare nominal is mapped. Specifically, an underspecified bare noun allows consideration of *singleton value sets* – defined over distinctly identifiable, hearer-old entities in the discourse context $\{x_i\}, \{x_2\}, .., \{x_h\}$ as well as a *non-singleton value set* that includes hearer-new entities satisfying the nominal's descriptive content $\{x_{h+1}, x_{h+2}, .., x_m\}$. This is in contrast to expressions associated with lexical uniqueness-based or *determined reference* presuppositions [14], like English *the* (or the bare nominal under an ambiguity view that is lexically specified for definiteness), which can only map to entities in singleton value sets[9]. The choice of the value set in the current view depends on the probability that a member of the set is present in the domain in which the bare nominal is interpreted: i.e., the *topic situation* of the sentence containing it (cf. [22]). To implement these ideas concretely, we can define a probabilistic computational model based on [17], which incorporates uncertainty about the identity of the situational domain restriction along with uncertainty about the value set in natural discourse scenarios:

[7] Even [11], which presents the most comprehensive discussion of the various uses of the (Hindi) bare nominal, indirectly separates the definite-like uses of the bare nominal from its existential uses. Bare nominals are ambiguous between definite and kind readings in her account, and existential readings are claimed to stem from the kind version of the bare nominal in episodic contexts.

[8] The proposed view most resembles the one suggested in [16], that the bare nominal instantiates only the indefinite meaning, which can be pragmatically strengthened to a definite meaning in appropriate contexts. Note however that we conceptualize the bare nominal as unspecified for (in)definiteness, not inherently indefinite.

[9] In [14], only those contexts are considered that make available a unique *salient* singleton set. In contexts containing more than just one comparably salient singleton set, a similar topicality-based inference process to the one described here for bare nominals may be needed to zero in on a particular choice of value set even in English. In such cases, we might expect to see gradient reference resolution behavior: see [25] for a presentation of such experimental data and related discussion.

(4) $P(x_j \mid \mathcal{RE}, \mathcal{C}) = \Sigma_{i=1}^{n}\ p(s_i)\ P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i)$

(5) $P(x_j \mid \mathcal{RE}, \mathcal{C}\ , s_i) = \frac{[\![Q(x_j)]\!]\ \mathrm{In}(x_j)(s_i)}{\Sigma_{k=1}^{m}\ [\![Q(x_k)]\!]\ \mathrm{In}(x_k)(s_i)}$

According to (4)-(5), when a bare nominal $\mathcal{RE}$ is uttered in a context $\mathcal{C}$, we jointly estimate how likely a particular situation s_i is to be the topic situation $p(s_i)$, and how likely a referent x_j belonging to a singleton or non-singleton value set is to be the intended referent $P(x_j \mid \mathcal{RE}, \mathcal{C}, s_i)$. The latter term is further composed of two components: (i) $[\![Q(x_j)]\!]$, denoting whether x_j satisfies the descriptive content Q of the nominal, and (ii) $0 \leq \mathrm{In}(x_j)(s_i) \leq 1$, denoting the extent to which we expect x_j at that point in the discourse to be present in the situation s_i. This expectation is estimated on the basis of common sense and world knowledge based reasoning. For example, in (1), it is *a priori* unclear that the the boy in the classroom was also present in yesterday's situation described in the subsequent sentence, so that the expectation of his being present in yesterday's situation is low. Other discourse-contextual factors play into estimating the probability $p(s_i)$ with which a situation s_i is the topic situation. Specifically, I adopt the view, following [22], that situational topicality is determined based on the Question Under Discussion (QUD; [21]), which in turn depends on certain discourse-structural cues such as whether the referent has been previously mentioned and the frequency of its mention, and grammatical cues such as word order and the presence of sentential adverbs. For instance, in (1), the topic situation for the sentence containing the bare nominal is set *yesterday*, indicated by the sentence-initial frame-setting adverb.

The present proposal resembles the one favored in [4] in its treatment of sentential situations as preferred domain restrictions in which the bare noun is interpreted. Indeed, the two accounts explain the data in (1)-(2) in much the same way – namely, by casting doubt on whether the hearer-old referent is a potential referent at all within the topic situation (while in the existing view, the hearer-old reading is categorically ruled out, the current account predicts it to lose to an interpretation where the bare nominal picks out a hearer-new referent). Notably, the underspecification-based proposal has an advantage in accounting for (3). The increased context around the mouse introduced in the first sentence in (3) increases the topic potential of the minimal situation surrounding this mouse, making it a viable candidate in this model to be the intended referent despite conflicting cues from the sentence-initial adverbial about the identity of the sentence situation. The gradient prediction made within the current proposal also aligns with the observation (for Mandarin) in [12], that "we are talking about preferences, not judgments of absolute (un)grammaticality".[10]

5 Concluding Remarks

This paper presents an alternative framework for analyzing the crosslinguistic anaphoric variability of bare nominals that goes against the grain of existing proposals, wherein the bare nominal is underspecified for (in)definiteness readings rather than ambiguous. I have argued that such a framework opens up a new analytical strategy in which existential interpretations of the bare nominals compete directly with the definite-like interpretations – a strategy that is categorically ruled out in the absence of underspecification. More broadly, this proposal questions the validity of restricting our study of a linguistic form to a subset of its interpretations solely

[10] A question that arises is whether underspecification is in fact necessary to account for (3), or if it suffices to make the notion of *situational familiarity* in Bremmers et al.'s proposal a probabilistic one while otherwise retaining the ambiguity assumption. Notice however that the uncertainty in the referent's identity relies crucially on the simultaneous consideration of alternative, non-singleton value sets in the current model, which is independent of the uncertainty in the topic situation itself.

on the basis of a lexical distinction that exists in another language. The proposed model of reference resolution in an underspecified bare nominal can be used to make concrete predictions about speakers' behaviors in specific contexts, and therefore adds to the space of hypotheses to consider in systematic future empirical work, for which the time is now ripe. While the discussion in this paper has collapsed over bare nominals across *article-less* languages, future work should also seek to evaluate finer-grained language-specific distinctions among these items.

There are other crosslinguistic implications of an underspecified bare nominal. The idea that bare nominals are not lexically specified for (in)definiteness makes salient the possibility that the meanings denoted by *a* and *the* in English do not represent deep universal semantic distinctions. Instead, it may be that they contribute more peripheral and potentially defeasible meanings in specific languages (e.g., as presuppositions or implicatures rather than at-issue entailments; cf. [8]). In other words, it would not be surprising under this picture to find an overlap between the distributions of *a* vs. *the*, given that at least some of the meanings they convey are contributed by the same lexical item in other languages. Some such cases have already been noted in the literature, where *the* is interchangeable with *a* in denoting a non-uniquely identifiable, exceptionally narrow-scoping meanings: for example, in describing culturally stereotypical activities such as in "go to the hospital" or "listen to the radio" (weak definites: [5], or even *good-enough descriptions* like "the house next door": [20]). The idea that the semantic contribution specific to the is weaker in the sense of being more peripheral/defeasible presents a new strategy for analyzing weak definites – one where this contribution is either suppressed or trivially satisfied, but where the nominal otherwise retains its at-issue meaning across weak and regular uses.

References

[1] Dorothy Ahn. Definite and demonstrative descriptions: a micro-typology. *GLOW in Asia 2017 Proceedings*, 2017.

[2] Dorothy Ahn. *THAT thesis: A competition mechanism for anaphoric expressions.* Ph.D. dissertation, Harvard University, 2019.

[3] Ruby Arkoh and Lisa Matthewson. A familiar definite article in Akan. *Lingua*, 123:1–30, 2013.

[4] David Bremmers, Jianan Liu, Martijn van der Klis, and Bert Le Bruyn. Translation mining: definiteness across languages (a reply to jenks 2018). *Linguistic Inquiry*, 53(4):735–752, 2022.

[5] Greg Carlson and Rachel Sussman. Seemingly indefinite definites. *Linguistic evidence: Empirical, theoretical, and computational perspectives*, 85:71–85, 2005.

[6] Gennaro Chierchia. Reference to kinds across language. *Natural language semantics*, 6(4):339–405, 1998.

[7] James N Collins. Definiteness determined by syntax. *Natural Language & Linguistic Theory*, 37(4):1367–1420, 2019.

[8] Elizabeth Coppock and David Beaver. Definiteness and determinacy. *Linguistics and Philosophy*, 38(5):377–435, 2015.

[9] Veneeta Dayal. The singular-plural distinction in Hindi generics. In *Semantics and Linguistic Theory*, volume 2, pages 39–58, 1992.

[10] Veneeta Dayal. Bare NPs, reference to kinds, and incorporation. In *Semantics and Linguistic Theory*, volume 9, pages 34–51, 1999.

[11] Veneeta Dayal. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, 27(4):393–450, 2004.

[12] Veneeta Dayal and Li Julie Jiang. The puzzle of anaphoric bare nouns in mandarin: A counterpoint to index! *Linguistic Inquiry*, pages 1–42, 2020.

[13] Amy Rose Deal and Julia Nee. Bare nouns, number, and definiteness in Teotitlán del Valle Zapotec. In *Proceedings of Sinn und Bedeutung*, volume 21, pages 317–334, 2018.

[14] Donka F Farkas. Specificity distinctions. *Journal of semantics*, 19(3):213–243, 2002.

[15] Irene Heim. Artikel und definitheit (articles and definiteness). In A. von Stechow and D. Wunderlich, editors, *Semantics: An International Handbook of Contemporary Research*. Berlin: de Gruyter, 1991.

[16] Irene Heim. Definiteness and indefiniteness. *Semantics: Noun Phrases and Verb Phrases*, 2011.

[17] Daphna Heller, Christopher Parisien, and Suzanne Stevenson. Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, 149:104 – 120, 2016.

[18] Peter Jenks. Two kinds of definites in numeral classifier languages. In *Semantics and Linguistic Theory*, volume 25, pages 103–124, 2015.

[19] Peter Jenks. Articulated definiteness without articles. *Linguistic Inquiry*, 49:501–536, 2018.

[20] Ann Reed. Good-enough 'the'. In *Poster Presented at Conference:"Definiteness across Languages" Mexico City*, 2016.

[21] Craige Roberts. Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5:6–1, 2012.

[22] Florian Schwarz. *Two types of definites in natural language*. PhD thesis, University of Massachusetts, Amherst, 2009.

[23] Sadhwi Srinivas. *The semantics of (in)definiteness in bare vs. non-bare nominals: A study of Kannada and English*. PhD thesis, Johns Hopkins University, 2021.

[24] Sadhwi Srinivas and Kyle Rawlins. Anaphoric variability in Kannada bare nominals. In *Proceedings of Formal Approaches to South Asian Linguistics (FASAL)*, volume 10, in press, 2020(a).

[25] Sadhwi Srinivas, Kyle Rawlins, and Daphna Heller. Asymmetries between uniqueness and familiarity in the semantics of definite descriptions. Presented at SALT 30, Cornell University, 2020.

Contrafactives and Learnability*

David Strohmaier[1] and Simon Wimmer[2]

[1] Department of Computer Science and Technology
ALTA Institute
University of Cambridge
ds858@cam.ac.uk
[2] Institut für Philosophie und Politikwissenschaft
Technische Universität Dortmund
simon.wimmer@tu-dortmund.de

Abstract

Richard Holton has drawn attention to a new semantic universal, according to which (almost) no natural language has contrafactive attitude verbs. This semantic universal is part of an asymmetry between factive and contrafactive attitude verbs. Whilst factives are abundant, contrafactives are scarce. We propose that this asymmetry is partly due to a difference in learnability. The meaning of contrafactives is significantly harder to learn than that of factives. We tested our hypothesis by conducting a computational experiment using an artificial neural network. The results of this experiment support our hypothesis.

1 Introduction

Holton (2017) has drawn attention to a novel semantic universal, according to which (almost) no natural language has contrafactive attitude verbs. Contrafactives are the mirror image of factive attitude verbs, such as *know*, *remember*, *see*, and *regret*. Although both factives and contrafactives entail a belief, contrafactives differ from factives in presupposing the falsity, as opposed to truth, of their declarative complements. To illustrate, suppose we extend our language with the contrafactive *contra*. Now, whilst both *Dan knows that Maggie is dancing* and *Dan contras that Maggie is dancing* entail that Dan believes that Maggie is dancing, the former presupposes that it is raining, the latter that it is not. This difference in presuppositions between the factive and contrafactive surfaces in the following diagnostics.

1. # Umut knows that it's raining, but it isn't.
2. # Umut contras that it's raining, and it actually is.
3. (a) Does Eylem know that we've won?
   (b) Eylem doesn't know that we've won.
   → We've won.
4. (a) Does Eylem contra that we've won?
   (b) Eylem doesn't contra that we've won.
   → We haven't won.

*This paper reports on research supported by Cambridge University Press and Assessment, University of Cambridge. We thank the NVIDIA Corporation for the donation of the Titan X Pascal GPU used in this research. Simon Wimmer's work on this paper was supported by a postdoc stipend of the Fritz Thyssen Foundation. We thank audiences in Bochum, Dortmund, Essen, Tokyo, and Utrecht for discussion of related material. David Strohmaier designed and ran the computational experiment, Simon Wimmer did conceptual work to connect philosophical and linguistic discussions to the experiment.

The diagnostic in 1 and 2 shows that the inference to the truth/falsity of the verb's declarative complement cannot be cancelled; the diagnostic in 3 and 4 suggests that the inference projects through entailment-cancelling environments, such as question and negation.

An important further feature of a contrafactive, according to Holton, is that it is an "atomic propositional attitude verb" (2017, p.248).[1] Thus, the inference to the falsity of its declarative complement is not the result of a compositional method: that would make the target expression "molecular". This means that Anvari, Maldonado, and Soria Ruiz's (2019) Spanish *creerse* 'wrongly believe' (as well as Shatz et al.'s (2003) Puerto Rican Spanish *creerse*) does not count as a contrafactive. Although *creerse* performs much as *contra* on the above diagnostics (with the exception of what Anvari, Maldonado, and Soria Ruiz call "polarity reversal under negation"), *creerse* is built by adjoining the reflexive pronoun *se* to the non-factive verb *creer* 'believe'.[2] And, since adjoining the reflexive pronoun has similar effects in the case of *pensarse* and French *s'imagine*, Anvari, Maldonado, and Soria Ruiz (2019, p.72) suggest that the inference to the falsity of *creerse*'s declarative complement results, at least partially, from composing the meanings of *se* and *creer*. Hence, *creerse* is not atomic, and so not a contrafactive.

Holton (2017, pp.245-9, 262–4) considered several apparent counterexamples to his universal found in non-Indo-European languages. However, none is a genuine contrafactive. For instance, the Mandarin verb *yǐwéi*, glossed by Lee, Olson, and Torrance (1999) and Cheung, Chen, and Yeung (2009) as 'believe wrongly', has been found to carry a post-supposition that the reported belief must not be added to the common ground, rather than a presupposition that its declarative complement is false (Glass, 2022): the inference *yǐwéi* triggers can be cancelled and does not project through negation. Likewise, the inferences triggered by Shatz et al.'s (2003) Turkish belief verbs *san* and *zannet* can be cancelled, too.[3]

That no natural language has contrafactives raises the question: why do natural languages universally have factives like *know* (Goddard, 2010; Haspelmath and Tadmor, 2009), but universally lack contrafactives? Importantly, the issue here would remain, even if some counterexamples to Holton's universal were eventually found. For even if there were some contrafactives, an asymmetry between factives and contrafactives would persist: factives would be abundant, contrafactives scarce.

Our aim here is to uncover one reason for the asymmetry between factives and contrafactives. Drawing on recent discussions of other semantic universals, like the veridical uniformity universal for responsive verbs (Steinert-Threlkeld, 2019), the conservativity, monotonicity, and quantity universals for determiners (Steinert-Threlkeld and Szymanik, 2019), and the convexity universal for color terms (Steinert-Threlkeld and Szymanik, 2020), we explore the hypothesis that the asymmetry between factives and contrafactives arises partly because the meaning of a contrafactive is harder to learn than that of a factive. Our hypothesis is inspired by the intuitive idea that languages have words for meanings that are easier to acquire and use compositional methods to express meanings that are harder to learn (Steinert-Threlkeld and Szymanik, 2019, p.4). We tested our hypothesis by conducting a computational experiment using an artificial neural network. As we explain below, the results of the experiment support our hypothesis.

[1] Holton adopts two further necessary conditions an expression must satisfy in order to count as a contrafactive. In parallel with *know*, he would regard *contra* as a mental state verb and as responsive (embedding both declarative and interrogative complements). For present purposes, however, we set these conditions aside. For one, we take the question of why no natural language has a verb with the features noted in the text to be of independent interest. For another, we expect the work we present here to also go some way toward addressing why no natural language has a verb that satisfies all of Holton's conditions.

[2] By contrast with factives and contrafactives, a non-factive, like *believe* or *think*, triggers neither an uncancellable inference to the truth/falsity of its declarative complement nor an inference to truth/falsity that projects through entailment-cancelling environments.

[3] We are grateful to Dilara Malkoc for discussion of the Turkish data.

2 Hypothesis

Before we turn to our computational experiment, we want to provide an intuitive motivation for our claim that factives are easier to learn than contrafactives. This motivation is inspired by Phillips and Norby's (2021) work on differences between factive and non-factive mental state attribution (see also Nagel, 2017; Phillips et al., 2020).

Suppose a speaker utters the factive attitude ascription *Dan knows that Maggie is dancing.* Since factives presuppose the truth of their declarative complements, this utterance commits the speaker to it being true that Maggie is dancing. Further, since factives entail a belief, the ascription entails that its subject, Dan, is also committed to it being true that Maggie is dancing. For this reason, the factive attitude ascription represents the ways the speaker and the subject of the ascription take the world to be as converging. Put in terms of Phillips and Norby's (2021) map analogy, the speaker's map of the world and the map they attribute share certain parts; our speaker simply copy-pastes these parts from their own map onto the other.

The ascription of non-factive attitudes is significantly less constrained. To see this, suppose a speaker utters the non-factive attitude ascription *Dan believes that Maggie is dancing.* Since non-factives do not presuppose the truth of their declarative complements, this utterance does not commit our speaker to it being true that Maggie is dancing. However, the ascription does entail that its subject, Dan, is committed to it being true that Maggie is dancing. For this reason, the non-factive attitude ascription by itself leaves open whether the ways the speaker and the subject of the ascription take the world to be converge or diverge. Thus, the speaker cannot simply copy-paste parts of their own map of the world onto the map they attribute. In this sense, their own take on what the world is like does not constrain, and so simplify, their ascription of a non-factive attitude to another person.

We can extend the point from non-factives to contrafactives. Suppose a speaker utters *Dan contras that Maggie is dancing.* Since contrafactives presuppose the falsity of their declarative complements, this utterance commits our speaker to it being false that Maggie is dancing. Yet the ascription entails that its subject, Dan, is committed to it being true that Maggie is dancing. Given this, the contrafactive attitude ascription represents the ways the speaker and the subject of the ascription take the world to be as diverging. And so, the speaker's map of the world and the map they attribute have incompatible parts. Consequently, they cannot copy-paste parts of their own map of the world onto the map they attribute. As for non-factives, the speaker's own take on what the world is like does not constrain, and so simplify, their contrafactive attitude ascriptions in the way in which it constrains their factive attitude ascriptions.

Importantly, whether a speaker's own take on what the world is like constrains, and thus simplifies, their ascription of factive and contrafactive attitudes is arguably (at least partly) due to the meaning of factives and contrafactives. This, however, leads us to expect the meaning of a contrafactive to be harder to acquire than that of a factive.

Building on Phillips and Norby (2021), we also expect the meaning of a contrafactive to be slightly easier to learn than that of a non-factive. For there is a sense in which a speaker's take on what is the case does constrain, even if only slightly, their contrafactive, but not their non-factive, attitude ascriptions. If, say, our speaker takes Maggie to be dancing, they cannot consistently claim that Dan contras that Maggie is dancing, but can consistently ascribe a belief to that effect. So, the speaker's take on what is the case rules out a contrafactive attitude ascription for them, but leaves open the corresponding non-factive attitude ascription. In this sense, information about the way the world is only contributes noise to the ascription of non-factive attitudes; noise one must learn to ignore. As we note below, our experimental results also support the claim that contrafactives are slightly easier to learn than non-factives.

The result that the meaning of a non-factive is harder to acquire than that of a contrafactive can seem puzzling, given that non-factives like *think* are universal across natural languages (Goddard, 2010). However, the added difficulty of acquiring the meaning of a non-factive just described does not entail that it is less common. Other factors that our current experiment does not model can make a contrafactive *overall* harder to learn than a non-factive. For instance, on the pragmatic syntactic bootstrapping model of how infants acquire attitude verb meanings (Hacquard and Lidz, 2022), the meaning of non-factive *think* is partly inferred from the parallel between the use of *I think P* as an indirect assertion and the primary use of *P* as an assertion, and the meaning of factive *know* is partly inferred from the parallel between the use of *Do you know Q?* as an indirect question and the primary use of *Q?* as a question. Yet unlike in the case of factives and non-factives, no such parallels would hold for a contrafactive. One cannot use *Dan contras that Maggie is dancing*, say, as an indirect assertion that Maggie is dancing. Thus, use of a contrafactive attitude ascription would not match the primary use of its complement. And so, we expect pragmatic syntactic factors to make it harder to acquire contrafactives than non-factives and factives. We leave it to future work to explore how this pragmatic syntactic difference between factive, non-factive, and contrafactive attitude ascriptions and the differences suggested by our computational experiment combine to explain the difference in frequency of factive, non-factive, and contrafactive attitude verbs in natural languages.

3 Experiment and results

To test our expectation that the meaning of a contrafactive is harder to acquire than that of a factive, we conducted a computational experiment using an artificial neural network, specifically a Transformer encoder.[4] This network was trained to predict the truth value of factive, non-factive, and contrafactive attitude ascriptions, given a representation of a small world and a representation of the small world as the attitude holder takes it to be (which may or may not be accurate).[5] The network's predictions were expressed in a probability within [0,1] that the target ascriptions are true. The artificial language in which these ascriptions were formulated and which the neural network learned can be interpreted as a fragment that describes propositions about the relative locations of two objects to each other plus the attitude taken towards these propositions. The small world can be conceived of as a 3-by-3 grid containing 3 objects. All objects differ in shape and they sometimes differ in colour. A typical statement in the artificial language can be glossed as *contra red triangle above blue square*, so long as we bear in mind that the network lacks any real world knowledge about triangles, squares, etc.

To encode this artificial language as well as the mind and world representations, we used a Transformer encoder from the pyTorch library. Transformers, based upon the so-called attention mechanism that allows contextualised processing of word information, are the foundation of current state-of-the-art results in natural language processing (Vaswani et al., 2017; Devlin et al., 2019; Rogers, Kovaleva, and Rumshisky, 2020). Our Transformer encoder used position embeddings and sequence embeddings to encode word order and distinguish the three types of input (attitude ascription, world representation, mind representation).

Generally speaking, the results of our experiment show the Transformer-encoder to perform better on factives than contrafactives. While the performance on non-factives was even worse, this is to be expected both from our intuitive motivation and the architecture of our network.

[4] The code of the model and further information are available on David Strohmaier's (2022) GitHub.

[5] Since our intuitive motivation did not touch on ascriptions in entailment-cancelling environments, we did not train the network to handle such ascriptions. Thus, our network does not model the presupposition projection of factive and contrafactive attitude ascriptions. We plan to fill this gap in a follow-up experiment.

Our network always processes both a mind and a world representation, although the latter only contributes noise in the case of non-factives. We suspect that, as humans learn non-factives, they develop a better input-gating mechanism than our model, which would increase their performance on non-factives. Below we focus on results that bear on the comparison between factives and contrafactives.

We evaluated 51 hyperparameter settings in an initial search.[6] Of these 18 performed below 60% accuracy and 27 exceeded 90% accuracy, suggesting that the network generally is able to learn our attitude verbs. In all except 4 of the 51 settings, the accuracy was higher for the factive than the contrafactive (but not significantly so, according to a Kruskal-Wallis test). We take this to suggest that the difference is due to the neural architecture rather than specific hyperparameter settings.

kind	accuracy	MAE
contrafactive	97.6%	0.0296
factive	97.8%	0.0255
non-factive	96.8%	0.0368

Table 1: Accuracy and MAE

The setting which performed best in the hyperparameter search, i.e. the one with the highest overall accuracy, was then applied to a hold-out test set. The results on this test set once again showed higher performance for the factives than the contrafactives. The difference in accuracy was small (0.2 percentage point, see table 1), because the model was trained on such a large sample of data (633981 examples) that it was able to successfully learn all attitude verbs.

Looking at accuracy, however, discards some information, since for an ascription that is true (not true), a prediction (not) above 0.5 is treated as accurate. By contrast, mean absolute error also considers how far the prediction strayed from the correct values of 0 (not true) and 1 (true). The differences for the mean absolute error are still small, but more striking (see table 1). A permutation significance test (resamples=9999) shows that the error is significantly larger for contrafactives than factives ($p < 0.01$), see figure 1. The training for the factive also proceeded faster than for the contrafactive, see figure 2, providing further support for a difference in how hard it is to learn the meaning of a contrafactive as opposed to factive. To give some numbers for intuition, after 100000 training examples the average loss for the factive is 0.39, while the average loss for the contrafactive is 0.54.

Post-experimental analysis suggests that many of the remaining errors by the network have the following source: The network struggles with evaluating a sentence like *contra red triangle above blue square* if the world or mind representation contains a red triangle next to a blue square, rather than one of the objects being missing altogether. The network was paying excessive attention to whether objects named in the artificial language sentences were present, ignoring whether the described relationship between the objects held. Put differently, the network struggles with reading the spatial relations from the linear enumeration of the 3-by-3 grid's 9 cells. This can be interpreted as a difficulty of dealing with word order, which is well-documented for Transformer models in the NLP literature (e.g. Pham et al., 2021).

To complicate the situation further, the target ascriptions differ in how their truth depends on the presence of objects. Notably, the factive attitude ascription can be true only if both objects mentioned (e.g. red triangle and blue square) are present in the grid, while the contrafactive ascription can be true regardless of whether both, only one, or neither of the two

[6] A list of available hyperparameters can be found in our online appendix on GitHub (Strohmaier, 2022).

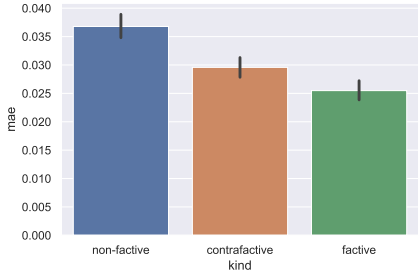


Figure 1: Mean absolute error on test set

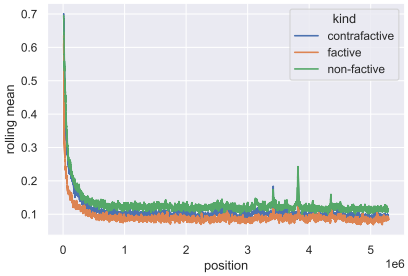


Figure 2: Rolling loss smoothed over 10000 instances during training

objects are present.[7] Given these findings, we are conducting a follow-up experiment in which the artificial language the neural network learns is simpler. The language we use here is interpretable as a fragment that merely describes a primitive, non-decomposable proposition and the attitude taken toward that proposition.

Our computational experiment improves on similar ones conducted by Steinert-Threlkeld (2019) and Steinert-Threlkeld and Szymanik (2019; 2020) in a number of ways. First, we report results based on a larger range of hyperparameters (e.g. training epochs, learning rate, etc.). For example, Steinert-Threlkeld and Szymanik (2019) only report the results in the case of two layers of LSTM cells and a hidden dimensionality of 12. We have explored 51 hyperparameter settings which can vary both in dimensionality and number of layers, among other hyperparameters, and can report that for all but four the accuracy for factives is higher than for contrafactives. This provides a better sense of the robustness of our experimental results. Second, while the cited research used feed-forward neural networks and LSTMs, we switched to the more advanced Transformer-architecture. Recent results suggest that, despite not being originally designed for cognitive plausibility, Transformer-based networks show greater convergence with human processing than other approaches (e.g. Caucheteux and King, 2022; Schrimpf et al., 2021). Given this, the results of our computational experiment likely reflect learnability for human language learners more closely than previous work.

4 Conclusion

Factives are abundant, contrafactives scarce in natural languages. We suggested that this asymmetry is partly due to a difference in learnability: the meaning of a contrafactive is significantly harder to learn than that of a factive. To support our suggestion, we reported the results of a computational experiment.

In closing, let us emphasize the scope of our discussion. We take our computational experiment to highlight one reason for the difference in frequency between factives and contrafactives in natural languages. But, most likely, this is not the only reason. We mentioned one other likely reason in section 2: the pragmatic syntactic difference between factive and contrafactive attitude ascriptions. We plan to explore this and other reasons in future work.

[7] This corresponds to the constraint on factive, but not contrafactive, attitude ascription noted earlier: the former is true only given a perfect match between world and mind, the latter can be true so long as there is no perfect match. So, there are many more ways for the latter to be true than for the former.

References

Anvari, Amir, Mora Maldonado, and Andrés Soria Ruiz (2019). "The puzzle of Reflexive Belief Construction in Spanish". In: *Proceedings of Sinn und Bedeutung* 23.1, pp. 57–74. DOI: 10.18148/sub/2019.v23i1.503. URL: https://ojs.ub.uni-konstanz.de/sub/index.php/sub/article/view/503.

Caucheteux, Charlotte and Jean-Rémi King (2022). "Brains and algorithms partially converge in natural language processing". In: *Communications Biology* 5.1, p. 134. DOI: 10.1038/s42003-022-03036-1. URL: https://www.nature.com/articles/s42003-022-03036-1 (visited on 08/15/2022).

Cheung, Him, Hsuan-Chih Chen, and William Yeung (2009). "Relations between mental verb and false belief understanding in Cantonese-speaking children". In: *Journal of Experimental Child Psychology* 104.2, pp. 141–155. DOI: 10.1016/j.jecp.2009.05.004. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022096509001155 (visited on 11/03/2021).

Devlin, Jacob et al. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. DOI: 10.48550/arXiv.1810.04805. URL: http://arxiv.org/abs/1810.04805 (visited on 11/27/2022).

Glass, Lelia (2022). "The Negatively Biased Mandarin Belief Verb yǐwéi*". In: *Studia Linguistica* n/a.n/a. DOI: 10.1111/stul.12202. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/stul.12202 (visited on 08/15/2022).

Goddard, Cliff (2010). "Universals and Variation in the Lexicon of Mental State Concepts". In: *Words and the Mind: How words capture human experience*. Oxford: Oxford University Press. URL: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195311129.001.0001/acprof-9780195311129-chapter-5 (visited on 03/11/2019).

Hacquard, Valentine and Jeffrey Lidz (2022). "On the Acquisition of Attitude Verbs". In: *Annual Review of Linguistics* 8.1, pp. 193–212. DOI: 10.1146/annurev-linguistics-032521-053009. URL: https://www.annualreviews.org/doi/10.1146/annurev-linguistics-032521-053009 (visited on 05/11/2022).

World Loanword Database (WoLD) (2009). Tech. rep. München: Max Planck digital library. URL: http://wold.livingsources.org/ (visited on 03/28/2022).

Holton, Richard (2017). "I—Facts, Factives, and Contrafactives". In: *Aristotelian Society Supplementary Volume* 91.1, pp. 245–266. DOI: 10.1093/arisup/akx003. URL: https://0-academic-oup-com.pugwash.lib.warwick.ac.uk/aristoteliansupp/article/91/1/245/3897122 (visited on 11/06/2018).

Lee, Kang, David R. Olson, and Nancy Torrance (1999). "Chinese children's understanding of false beliefs: the role of language". In: *Journal of Child Language* 26.1, pp. 1–21. DOI: 10.1017/S0305000998003626. URL: https://www.cambridge.org/core/journals/journal-of-child-language/article/chinese-childrens-understanding-of-false-beliefs-the-role-of-language/B0A84B10F7113D94BB3330563EEB4248 (visited on 11/03/2021).

Nagel, Jennifer (2017). "Factive and nonfactive mental state attribution". In: *Mind & Language* 32.5, pp. 525–544. DOI: 10.1111/mila.12157. (Visited on 11/14/2017).

Pham, Thang et al. (2021). "Out of Order: How important is the sequential order of words in a sentence in Natural Language Understanding tasks?" In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Online: Association for Computational Linguistics, pp. 1145–1160. DOI: 10.18653/v1/2021.findings-acl.98. URL: https://aclanthology.org/2021.findings-acl.98 (visited on 11/27/2022).

Phillips, Jonathan and Aaron Norby (2021). "Factive theory of mind". In: *Mind & Language* 36.1, pp. 3–26. DOI: https://doi.org/10.1111/mila.12267. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/mila.12267 (visited on 02/19/2021).

Phillips, Jonathan et al. (2020). "Knowledge before Belief". In: *Behavioral and Brain Sciences*, pp. 1–37. DOI: 10.1017/S0140525X20000618. URL: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/knowledge-before-belief/B434EF04A3EA77018384EABEB4973994 (visited on 10/07/2020).

Rogers, Anna, Olga Kovaleva, and Anna Rumshisky (2020). "A Primer in BERTology: What We Know About How BERT Works". In: *Transactions of the Association for Computational Linguistics* 8, pp. 842–866. DOI: 10.1162/tacl_a_00349. URL: https://aclanthology.org/2020.tacl-1.54 (visited on 11/27/2022).

Schrimpf, Martin et al. (2021). "The neural architecture of language: Integrative modeling converges on predictive processing". In: *Proceedings of the National Academy of Sciences* 118.45, e2105646118. DOI: 10.1073/pnas.2105646118. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2105646118 (visited on 11/27/2022).

Shatz, Marilyn et al. (2003). "The influence of language and socioeconomic status on children's understanding of false belief". In: *Developmental Psychology* 39.4, pp. 717–729. DOI: 10.1037/0012-1649.39.4.717.

Steinert-Threlkeld, Shane (2019). "An Explanation of the Veridical Uniformity Universal". In: *Journal of Semantics*. DOI: 10.1093/jos/ffz019. URL: https://0-academic-oup-com.pugwash.lib.warwick.ac.uk/jos/advance-article/doi/10.1093/jos/ffz019/5683663 (visited on 12/24/2019).

Steinert-Threlkeld, Shane and Jakub Szymanik (2019). "Learnability and semantic universals". In: *Semantics and Pragmatics* 12.0, p. 4. DOI: 10.3765/sp.12.4. URL: https://semprag.org/index.php/sp/article/view/sp.12.4 (visited on 11/18/2019).

— (2020). "Ease of learning explains semantic universals". In: *Cognition* 195, p. 104076. DOI: 10.1016/j.cognition.2019.104076. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010027719302495 (visited on 03/18/2022).

Strohmaier, David (2022). *Contrafactives: Exploration of a Grid World*. URL: https://github.com/dstrohmaier/contrafactives_grid_world (visited on 11/30/2022).

Vaswani, Ashish et al. (2017). "Attention is All you Need". In: *31st Conference on Neural Information Processing Systems*, pp. 1–11.

Morphological effects on indexical shift in Uyghur*

Ruoan Wang

Massachusetts Institute of Technology
rmwang@mit.edu

Abstract

Novel fieldwork on Uyghur shows indexical shift to be sensitive to *morphological form*. Isolated reduced indexicals always optionally shift (despite various manipulations to structural position). This is surprising for an operator-based theory, and calls for indexical-by-indexical specifications of shifting *à la* pronoun-based theories. However, this optionality disappears when full and reduced indexicals co-occur, suggesting that shiftability is determined by morphological makeup instead. Current theories of indexical shift are not form-sensitive and thus inadequate. I propose that a hierarchy of pronominal forms differentiating reduced from full indexicals, and stipulate that reduced indexicals can be uninterpretable (and thus invisible to an operator), but full ones cannot.

1 Theories of (Uyghur) indexical shift

Indexical shift is the phenomenon where the reference of an embedded indexical may shift *away* from its reference in the matrix context of utterance.

English is a non-indexical shift language. *I* can only refer to the speaker of the matrix utterance, never to the speaker of the reported utterance (1a). Shifting is only licensed in quotations (1b). In contrast, Uyghur (Turkic) exhibits indexical shift in finite clausal complements, where nominative-marked subjects shift. *mën* 'I' can only refer to the speaker of the reported utterance, Ahmet; not to the speaker of the matrix utterance (1).

(1) a. My father said **I** left. (=speaker, ≠speaker's father)
b. My father said, "**I** left". (≠speaker, =speaker's father)

(2) Ahmet [**mën** kät-tim] di-di.
Ahmet [1SG.NOM leave-PST.1SG] say-3PST
'Ahmet said 1SG (≠speaker, =Ahmet) left.' (Sudo & Shklovsky 2014: 383)

As the reference of indexicals was traditionally thought to be fixed to only the utterance context (e.g. Kaplan 1977), formalizing the mechanics of indexical shift has been greatly interesting for generative semantics. Current theories are broadly divided between two families, which are *operator-* or *pronoun-based*.

Operator-based theories propose a tight link between an indexical's height relative to an operator 🫥 and its shiftability (e.g. Anand & Nevins 2004, Anand 2006, Sudo & Shklovsky 2014, Deal 2020). A covert semantic operator 🫥 overwrites components of the context c with components from the index i, essentially overwriting c with i (3). This erases any information of the matrix speech act, so that all indexicals are interpreted relative to the reported speech act as a result.

(3) $[\![\ 🫥\text{XP}\ ]\!]^{c,i} = [\![\ \text{XP}\ ]\!]^{i,i}$

*Many thanks to Yasutada Sudo, my Uyghur consultant, Norvin Richards, and Amir Anvari for their helpful comments and insights in informing this work. All remaining errors are mine.

👻 is an independent lexical item, selectively introduced into the structure by attitude predicates like *say*. Whether 👻 is present is a matter of parametric variation. Not all attitude verbs introduce 👻 into their complement, and not all languages have 👻 in their lexicon. Since shiftability is tightly linked with height, an important prediction of such theories is that when 👻 is present, all indexicals within a clause must be evaluated with respect to the *same* context. This is known as *Shift Together* (Anand 2006: 100). The meanings of indexicals themselves are cross-linguistically uniform, since shifting only occurs within the scope of an independent context-shifting operator 👻 . The burden of shifting lies upon the position of 👻 .

A competing theory is *pronoun-based* (Schlenker 1999, 2003), which denies any link between an indexical's height and its shiftability. Rather, all attitude verbs are monstrous operators quantifying over contexts (4).

(4) $[\![\text{ say } _{<\alpha',\beta',\gamma'>} c_i \phi \,]\!]^{c,s} = 1$, iff for all c' compatible with the claim made by individual $[\![\alpha \,]\!]^{c,s}$ at time $[\![\beta \,]\!]^{c,s}$ in world $[\![\gamma \,]\!]^{c,s}$: $[\![\phi \,]\!]^{c,s[c_i \rightarrow c']} = 1$

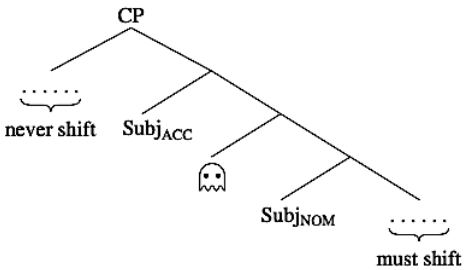
Each indexical carries its own lexical presupposition, whether they are bound by regular context pronouns c_i or a distinguished context pronoun c^*. c_i picks out reported contexts, resulting in shifted readings; c^* picks out the matrix utterance context, resulting in unshifted readings. Shift is defined on an indexical-by-indexical basis. Ambiguity is expected—indexicals have the choice to bind either c_i or c^*. The burden of shifting lies upon each indexical.

Previous work on Uyghur indexical shift has favored the operator-based theory (Sudo & Shklovsky (2014, henceforth S&S). Strikingly, Uyghur exhibits *partial indexical shift* which is sensitive to case of the embedded subject. When the subject of a finite embedded clause is *nominative* (5a), it must be shifted. Changing only the case of the embedded subject, so that it is *accusative* (5b), the indexical must be unshifted.

(5) a. Ahmet [**mën** kät-tim] di-di.
Ahmet [**1sg.nom** leave-1SG.PST] say-3PST
'Ahmet said I (≠speaker, =Ahmet) left.'
b. Ahmet [**mëni** kät-ti] di-di.
Ahmet [**1sg.acc** leave-3PST] say-3PST
'Ahmet said I (=speaker, ≠Ahmet) left.' (S&S 2014:383, 386)

S&S sandwich 👻 between embedded accusative and nominative subjects, as in (6). This configuration is responsible for partial indexical shift in Uyghur, and is supported by other diagnostics (involving long-distance *wh-* and NPI dependencies) showing that accusative embedded subjects are structurally higher than nominative ones. Notably, *Shift Together* falls out for free since 👻 is fixed.

(6)



Under a pronoun-based theory, the Uyghur partial indexical shift pattern are not so easy to capture. Since (6) applies to all accusative or nominative embedded subjects, regardless of

whether they are indexicals, it is hard to fix an individual indexical's shifting in a way that covaries systematically with case. Additionally, *Shift Together* must be independently stipulated.

In what follows, I present novel data from Uyghur which additionally considers the bound 1SG indexical *-im* (along with the previously investigated full indexicals *mën/mëni*). Bound indexicals in isolation are recalcitrantly ambiguous, *i.e.* exhibiting optional indexical shift, despite various manipulations to structural position (§2.1). Yet, this persistent ambiguity disappears with manipulations to morphological form (§2.2). Neither the operator- nor pronoun-based theory can capture both puzzles, since neither considers the morphological form of an indexical, which is the crucial factor here. I capture Uyghur's form-sensitive shifting by positing a hierarchy of pronominal forms (e.g. Patel-Grosz & Grosz 2017), where reduced indexicals can be uninterpretable, but full ones cannot (§3), making a connection to similar facts in Mishar Tatar and Austrian Bavarian.

2 Form-sensitive shifting in Uyghur

2.1 Manipulating structural positions

The bound indexical informing this puzzle is *-im*, the 1SG possessive suffix, where it is *ambiguous* between shifted and unshifted readings (7a). Ambiguity persists even after the embedded subject is accusative (7b). This is surprising for S&S' analysis in (6), as the case of the embedded subject directly affects shiftability, contradicting the paradigm of partial indexical shift.

(7) a. Ahmet [Dilyar-∅ imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-NOM test-1SG.POSS-ABL pass-3PST say-3PST
'Ahmet said Dilyar passed 1SG's **(=speaker, =Ahmet)** test.' (8:7)

b. Ahmet [Dilyar-ni imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-ACC test-1SG.POSS-ABL pass-3PST say-3PST
'Ahmet said Dilyar passed 1SG's **(=speaker, =Ahmet)** test.' (8:5)

The puzzle deepens as ambiguity persists after varied manipulations to the bound indexical's position. Recalling (6), an ACC embedded subject earmarks a position above 👻 (so indexicals above it must shift), and a NOM embedded subject a position below 👻 (so indexicals below it must not shift).

The first structural manipulation is to scramble an indexical above the 👻 operator, as (6) predicts that scrambling an indexical above ACC embedded subjects should force an unshifted reading. (8) exhibits scrambling of the indexical-containing phrase *periqlik oqughüch-im-din* 'different student of 1SG' above the accusative subject *herbir oqutqüchi-ni* 'every teacher-ACC'. Yet, the ambiguity of *-im* persists (albeit preferring an unshifted reading).

(8) Adil [periqlik oqughüch-**im**-din herbir oqutqüchi-ni päxirlini-du] di-di.
Adil different student-1SG.POSS-ABL every teacher-ACC proud-3NPST say-3PST
'A said every teacher is proud of a diff. student of 1SG's (=speaker, =?Tursun).' (7:31)

The second structural manipulation makes use of Principle A reconstruction. Consider a scenario where Tursun draws a picture of himself on Adil's desk. (9) displays scrambling of the phrase containing both the indexical *üm* and the anaphor *öz* (contained in *üstel-üm-diki öz-ning resim-i-ge* 'a picture of oneself on 1SG's desk'.)

(9) Adil [[üstel-**üm**-diki öz-ning resim-i-ge] Tursun-∅ siz-di] di-di.
Adil desk-1SG.POSS-REL self-GEN picture-3POSS-LOC Tursun-NOM draw-3PST say-3PST

'Adil said Tursun drew a picture of himself on 1SG's (=speaker, =Adil) desk.' (7:13)

According to Principle A, for the reflexive *öz-* to be bound by its antecedent *Tursun*, *öz-* must be lower than *Tursun* at LF. However, since (9) is grammatical, Principle A reconstruction has occurred. The anaphor *öz* is bound in its LF position, below *Tursun-∅* and hence below 𐤀 . Since the indexical is contained within the same phrase as the anaphor, it should also be interpreted at its lower LF position, yielding only a shifted reading. Yet, ambiguity is possible in (9). The failure of Principle A considerations to "un-shift" the indexical is also seen with another anaphor *birbir*, 'each other' (10).

(10) Adil [[sinip-**im**-diki birbiri-ning yaxshi köridighan oqughüchi-si-gha]
Adil class-1SG.POSS-REL each.other-GEN good see.PSTPART student-3POSS-DAT
Tursun we Norvin-∅ xet ewet-ti] di-di.
Tursun and Norvin-NOM letter send-3PST say-3PST
'Adil told me that Tursun and Norvin sent a letter to each other's favorite students in 1SG's (=speaker, =Adil) class.' (7:7)

The third structural manipulation is to make use of pragmatic reconstruction. Reconstruction can also be forced by pragmatic considerations (Fox 1995). Consider (11), which contains the ∀ phrase *hazir herbir* 'every girl' and the indexical-containing ∃ phrase *üch dost-üm bilen* 'with three of 1SG's friends'. A surface scope interpretation is pragmatically forced because mahjong is strictly a four-player game.

(11) hazir herbir qiz üch dost-**üm** bilen majang oyna-wati-du.
now every girl three friend-1SG.POSS with mahjong play-PRG-3NPST
'Every girl is playing mahjong with three of my friends now.' (5:7a)

Now, we embed (11) under 'say', creating the context for indexical shift, and also disrupt the pragmatically felicitous surface scope by scrambling. The result is (12):

(12) Adil [[üch dost-**üm** bilen] herbir qiz-**ni** majang oyna-wati-du] di-di.
Adil three friend-1SG.POSS with every girl-ACC mahjong play-PRG-3NPST say-3PST
'Adil said that every girl is playing mahjong with three of 1SG's (==speaker, =Adil) friends.' (5:8d, 7:23)

In addition to persistant ambiguity, the pragmatically plausible reading ($\forall > \exists$) is still the only one which holds after scrambling. This shows that the phrase containing the existential—and the indexical—is interpreted below the universal. Since the universal is an ACC embedded subject marking the position above 𐤀 , we see that scrambling above 𐤀 did not destroy ambiguity of the bound indexical *üm*.

The examples above were replicated across different predicates, scenarios, and sessions, so such ambiguity is neither example-specific nor session-specific. None of the three structural manipulations did not make ambiguity disappear, suggesting that shiftability and position are less tightly linked than operator-based theories assume. However, now we will see a factor which *does* make ambiguity disappear: co-occurrence with a full indexical.

2.2 Manipulating morphological form

Uyghur possession can be expressed via a fuller construction, where the possessive suffix *-im* is accompanied by a full genitive pronoun *mening*. Strikingly, addition of the full pronoun *mëning*

destroys the ambiguity that was so recalcitrant before. (13) shows this with a nominative subject; (14) the same with an accusative subject. This shows a lack of co-variance with case.

(13) a. Ahmet [Dilyar-∅ imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-NOM test-1SG.POSS-ABL pass-3PST say-3PST
'Ahmet said that Dilyar passed 1SG's (=speaker, =Ahmet) test.' (8:7)

b. Ahmet [Dilyar-∅ **mëning** imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-NOM 1SG.GEN test-1SG.POSS-ABL pass-3PST say-3PST
'Ahmet said that Dilyar passed 1SG's (≠speaker, =Ahmet) test.' (8:8)

(14) a. Ahmet [Dilyar-ni imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-ACC test-1SG.POSS-ABL pass-3PST say-3PST
'Ahmet said that Dilyar passed 1SG's (=speaker, =Ahmet) test.' (8:5)

b. Ahmet [Dilyar-ni **mëning** imtihan-**im**-din öt-ti] di-di.
Ahmet Dilyar-ACC 1SG.GEN test-1SG.POSS-ABL pass-3PST say-3PST
'Adil said that Dilyar passed 1SG's (=speaker, ≠Ahmet) test.' (8:6)

Yet, there *is* a correlation between case of the embedded subject and shiftability of indexicals underneath it, as per (6): indexicals under a NOM subject shift; all indexicals under a ACC one do not, swinging favor back to operator-based theories. This is especially because shifting coheres to *Shift Together* (15):

(15) a. Tursun [**mën** oqughüchi-**im**-din paxirlini-men] di-di.
Tursun 1SG.NOM student-1SG.POSS-ABL proud.of-1NPST say-3PST
'Tursun said that 1SG (≠speaker, =Tursun) am proud of 1SG's (≠speaker, =Tursun) student.' (8:13)

b. Tursun [**mëni** oqughüchi-**im**-din paxirlini-du] di-di.
Tursun 1SG.ACC student-1SG.POSS-ABL proud.of-3NPST say-3PST
'Tursun said that 1SG (=speaker, ≠Tursun) am proud of 1SG's (=speaker, ≠Tursun) student.' (8:14)

3 A proposal for form-sensitive shifting

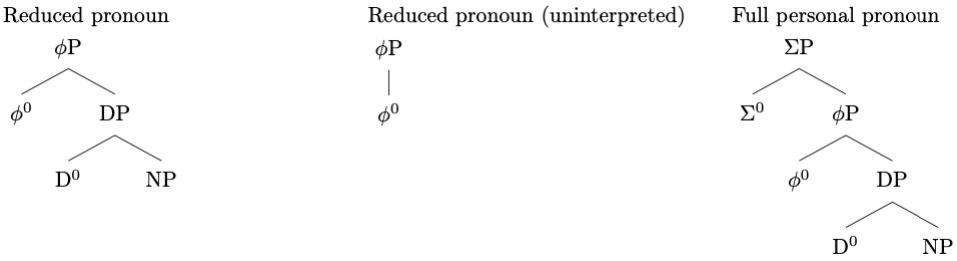
Neither the operator- nor pronoun-based theory explains the Uyghur facts. The operator-based theory, tightly linking position and shiftability, cannot explain why manipulation to structural position never destroys ambiguity, but manipulation to morphological form does. The pronoun-based theory, unlinking position from shiftability, cannot explain why the destruction of ambiguity is both *Shift Together*-adherent and S&S-adherent. Hence, I formalize a distinction between full vs. bound indexicals—within an operator-based theory—to capture the facts.

I adopt a hierarchy of pronominal forms (in the spirit of Patel-Grosz & Grosz (2017) and Cardinaletti & Starke (1999)), where reduced pronouns are weaker than full personal pronouns (with the strongest being a demonstrative pronoun). Here, "strength" correlates with the amount of functional material projected, as in (16)—ϕ encodes the purely formal phi-features, and Σ encodes features that forms pronouns into independent prosodic words. Another asymmetry: it is possible for reduced pronouns to be *uninterpretable* so that it lacks semantically interpreted material, but not possible for full pronouns, as Σ is not vacuous (presumably containing features pertaining to spell-out).

Crucially for our purposes, it is the uninterpretable flavor that is present in Uyghur reduced indexicals, rendering it invisible to semantic operators like 👻 explaining their recalcitrant am-

biguity in isolation (§2.1). Furthermore, recall that reduced indexicals' ambiguity disappears when co-occurring with a full one, in a fashion that coheres to *Shift Together* and Sudo & Shklovsky's analysis in (6) (§2.2). I stipulate that full pronouns must always be interpreted, so that 👻 must apply when present. Under an interpreted full pronoun, I further stipulate that reduced pronouns must be interpreted also. This could be due to an economy constraint where interpretability between two clausemate pronouns cannot differ; this correctly predicts the Uyghur pattern where the reduced pronoun "inherits" the shiftability of the full one.

(16)



Such a proposal would not be specific to the facts here, being further motivated by other links between the morphological makeup and interpretation. In Mishar Tatar (Turkic, Podobryeav 2014), covert vs. overt pronouns show an asymmetry in shiftability very similar to the Uyghur asymmetry here. In isolation, covert indexicals are ambiguous (17a), but this ambiguity disappears when it co-occurs with an overt one (17b).

(17) a. Alsu [**pro** kaja kit-te-m diep] at'-t?
Alsu 1SG.NOM where go.out-PST-1SG C say-PST
'Which place did Alsu say $1SG_{pro}$ (=speaker, =Alsu) went?'

b. Alsu [**pro** ber kajcan da **mija** bag-mas-mvn diep] bel-ä.
Alsu 1SG.NOM one when npcl 1SG.DAT look.at-NEG-1SG C know-ST.IPFV
'Alsu knows $1SG_{pro}$ (≠speaker, =Alsu) would never look at 1SG (=speaker, ≠Alsu).'
(Podobryeav 2014: 84, 86)

In Austrian Bavarian (Indo-European) dream reports, clitic vs. full pronouns show an asymmetry in interpretation: the clitic prefers to be read *de se* (18a), and the full pronoun *de re* (18b). The weakness of pronominal form correlates with being read *de se*, which is arguably an unmarked reading according to e.g. Pearson & Dery (2014). (Note that the Mishar covert pronoun in (17b), being shifted, is obligatorily read *de se* like the Bavarian clitic in (18a). This allows dream reports to make a connection to indexical shift.)

(18) a. I håb traamt, das 'e/i mei Nåchba bin und [das**'e** reich bin].
I have dreamed that $1SG_{cl/full}$ my neighbor am and that=**$1SG_{cl}$** rich am
'I dreamed that I am my neighbor and that $1SG_{cl}$ (=my neighbor) am rich.'

b. I håb traamt, das 'e/i mei Nåchba bin und [das **i** reich bin].
I have dreamed that $1SG_{cl/full}$ my neighbor am and that **$1SG_{full}$** rich am
'I dreamed that I am my neighbor and that $1SG_{full}$ (=actual speaker) am rich.'
(Patel-Grosz 2020: 563)

Of course, regardless of evidence from other languages, the stipulations made in this paper should also be supported by more Uyghur data: extending the data to cover 2nd person indexicals, for example, and testing if other non-possessive bound indexicals also display ambiguity. However, we hope to have evinced the need for a form-sensitive theory of indexical shift.

References

[1] Pranav Anand. *De de se*. PhD thesis, Massachusetts Institute of Technology, 2006.

[2] Pranav Anand and Andrew Nevins. Shifty operators in changing contexts. *Semantics and Linguistic Theory*, 14:20–37, 2004.

[3] Anna Cardinaletti and Michal Starke. The typology of structural deficiency: A case study of the three classes of pronouns. In H. van Riemsdijk, editor, *Clitics in the languages of Europe*, pages 145–233. Berlin: de Gruyter, 1999.

[4] Amy Rose Deal. *A theory of indexical shift: meaning, grammar, and crosslinguistic variation*, volume 82. MIT Press, 2020.

[5] Danny Fox. Economy and scope. *Natural Language Semantics*, 3(3):283–341, 1995.

[6] David Kaplan. Demonstratives. UCLA, Los Angeles, CA, 1977.

[7] Pritty Patel-Grosz. Pronominal typology and the *de se/de re* distinction. *Linguistics and Philosophy*, 43(5):537–587, 2020.

[8] Pritty Patel-Grosz and Patrick G Grosz. Revisiting pronominal typology. *Linguistic Inquiry*, 48(2):259–297, 2017.

[9] Hazel Pearson and Jeruen Dery. Dreaming 'de re' and 'de se': Experimental evidence for the Oneiric Reference Constraint. In *Proceedings of Sinn und Bedeutung*, volume 18, pages 322–339, 2014.

[10] Alexander Podobryaev. *Persons, imposters, and monsters*. PhD thesis, Massachusetts Institute of Technology, 2014.

[11] Philippe Schlenker. *Propositional attitudes and indexicality: a cross categorial approach*. PhD thesis, Massachusetts Institute of Technology, 1999.

[12] Philippe Schlenker. A plea for monsters. *Linguistics and Philosophy*, 26(1):29–120, 2003.

[13] Kirill Shklovsky and Yasutada Sudo. The syntax of monsters. *Linguistic Inquiry*, 45(3):381–402, 2014.

Revisiting presuppositional accounts of homogeneity *

Jad Wehbe

Massachusetts Institute of Technology
jadwehbe@mit.edu

Abstract

Homogeneity effects with definite plurals have been argued to constitute a phenomena that is distinct from other types of truth-value gaps [Kri15, KS21]. This paper makes the case for a return to early accounts of homogeneity that treat it as a presupposition. I provide empirical support for this view by showing that homogeneity is sensitive to constraints on presupposition accommodation. I argue that connectedness [EC21] mediates the distribution of local accommodation and that the differences between standard presuppositions and homogeneity are due to the fact that only the latter is non-connected (c.f. [Fox18], which argues for an alternative based on plausibility considerations).

1 Introduction

Definite plurals give rise to an intuitive truth-value gap known as *homogeneity*. Consider (1): neither (1-a) nor its negated counterpart in (1-b) are true in a scenario where Mary read only some of the books. Early accounts of homogeneity treated it as a presupposition [Sch93, Löb00, Gaj05], illustrated in (2)[1], but this assumption has recently been challenged.

(1) a. Mary read the books is **true iff** Mary read **all** the books.
b. Mary didn't read the books is **true iff** Mary read **none** of the books.
c. **Neither is true** iff Mary read only some of the books.

(2) **Homogeneity presupposition:** Given a distributive predicate P and a plural individual x, P(x) is defined iff $\forall y \leq_{\text{AT}} x : P(y) = 1 \vee \forall y \leq_{\text{AT}} x : P(y) = 0$

A major challenge for the presuppositional approach are differences in projection between homogeneity and standard presuppositions [Spe13, Kri15]. For example, it is well-known that presuppositions project from polar questions and from the antecedents of conditionals, but homogeneity does not seem to pattern with presuppositions in these environments (3).

(3) a. If Mary read the books, she passed the exam.
$\not\rightsquigarrow$ Mary read all or none of the books.
b. Did Mary read the books?
$\not\rightsquigarrow$ Mary read all or none of the books.

Furthermore, homogeneity violations aren't perceived in the same way as presupposition violations. Given a context where the presupposition is not entailed by the common ground such as (4-a), the addressee can object by responding with something like *Wait a minute! I didn't even know that Mary used to smoke* [vF04]. On the other hand, (4-b) does not impose this requirement on the common ground and therefore does not license the same kind of response.

*Thanks to Amir Anvari, Athulya Aravind, Itai Bassi, Omri Doron, Kai von Fintel, Danny Fox, Martin Hackl, Filipe Hisao Kobayashi, Keely New and the audiences of LFRG and HNM1 for helpful feedback.

[1] Throughout the paper, I will focus only on homogeneity with distributive predicates. Additionally, I abstract away from non-maximal interpretations of definite plurals. Both homogeneity with non-distributive predicates and the role of non-maximality should be considered more thoroughly in future work.

(4) a. **Context:** B does not know that Mary used to smoke.
A to B: Mary stopped smoking.
b. **Context:** B doesn't know whether Mary read all, some or none of the books.
A to B: Mary read the books.

In this paper, I will argue that homogeneity is a presupposition, despite these differences. In section 2, I show that homogeneity is sensitive to constraints on presupposition accommodation. In section 3, I argue that obligatory local accommodation with homogeneity in cases like (3-a) and (3-b) can explain the lack of projection and propose that these differences in projection follow from the fact that homogeneity gives rise to non-connected propositions (in the sense of [EC21]), while standard presuppositions are generally connected.

2 Constraints on accommodation

Heim observes that questions can't be answered with an accommodated presupposition [Hei15]. Consider (5): the only difference between (5-b) and (5-c) is that part of the information in (5-c) is presuppositional. The presupposition of (5-c) (*That someone adopted the Lab*) is not met in the context, but given the availability of global accommodation, this alone can't explain the infelicity. Heim argues that examples like (5-c) are infelicitous because accommodating the presupposition settles A's question, making the assertion trivial with respect to the QUD.

(5) **Context:** A is visiting a dog shelter and is interested in adopting a Lab that she saw the week before when passing by the shelter. She comes back and asks an employee:
a. A: Can I adopt the Labrador?
b. B: Someone from NY adopted the Lab.
c. B': #It is someone from NY who adopted the Lab. [AFH22] (slightly modified)

We can therefore make explicit the constraint on accommodation in (6), which follows from an asymmetry between the pragmatic status of presuppositions and assertions [DW22]. Taking PAI as a diagnostic for presuppositions, we can apply it to homogeneity. If homogeneity is a presupposition, we expect definite plurals to be infelicitous given a common ground where accommodating homogeneity makes the assertion uninformative.

(6) **Post-Accommodation Informativity (PAI):** A sentence S_p (presupposing p) can be uttered felicitously only if S_p is informative w.r.t. the common ground and QUD after presupposition accommodation. [DW22]

This prediction is borne out in (7-a). Given a common ground that entails that Jane read at least some of the blue books, accommodating the homogeneity presupposition (*She read none or all of the blue books*) makes the assertion (*She read the blue books*) trivial, regardless of the QUD. Therefore PAI correctly predicts the infelicity of (7-a). Compare with the counterpart with *all* which lacks homogeneity and is felicitous (7-b).

(7) **Context:** Mary and Bill were observing which books Jane is reading among a collection of blue and red books. They saw Jane read 3 out of the 6 blue books, but then Mary had to leave. Bill stayed and saw Jane finish all the blue books. Mary comes back to the room and is wondering which books Jane ended up reading. Bill reports:
a. B to M: # She read the blue books.
b. B to M: She read all the blue books.

In order to get a sharper contrast, consider the example in (8) (Bassi, p.c., atribb. to Schwarzschild), where the common ground that would license accommodation is not conceivable. Since it is common knowledge that Bill can't be married to all of his teammates, accommodating the homogeneity presupposition makes the assertion in (8-a) trivial, violating PAI.

(8) a. # Bill isn't married to his teammates.
b. Bill isn't married to any of his teammates.

In joint work with Omri Doron [DW22], we argue for a version of this constraint that applies also to local accommodation. Looking at conditional antecedents, a conditional is only felicitous if the common ground doesn't entail its antecedent. We argue that this requirement has to be evaluated after local accommodation: if the antecedent is trivial after accommodating its presupposition, local accommodation is not allowed. This is illustrated by the unavailability of local accommodation in cases like (9): given that it is common ground that Mary doesn't smoke now, once we accommodate the presupposition of the antecedent (*Mary used to smoke*), the antecedent proposition (*Mary stopped smoking*) is trivially true, violating PAI.

(9) I know that Mary doesn't smoke now. # I don't know if she used to smoke, but if she stopped smoking, she must be very brave.

Homogeneity is also sensitive to this constraint when the definite plural is in the antecedent of a conditional (10). Note that this is an environment from which homogeneity doesn't project and therefore has to be locally accommodated if it is a presupposition (see section 3). In (10), given a common ground where Jane read at least some of semantics books, accommodating the homogeneity presupposition makes the antecedent trivially true, violating the constraint.

(10) **Context:** A and B were looking at Jane, and they saw that she has read 3 out of the 6 semantics books on the reading list, but they're not sure if she read the rest of them.
a. A to B: # If she read the semantics books, she will pass the exam.
b. A to B: If she read all of the semantics books, she will pass the exam.

We therefore see that homogeneity is sensitive to the same constraints on accommodation as presuppositions, even in environments where a homogeneity presupposition would have to be locally accommodated. This parallel between homogeneity and presuppositions provides evidence that homogeneity is in fact a presupposition.

3 Connectedness and homogeneity

Starting with the assumption that homogeneity is a presupposition, I will show in this section that the projection patterns with homogeneity can be explained we constrain the distribution of local accommodation in the right way, building on [Fox18].

I assume that sentences with presuppositions denote trivalent propositions and that local accommodation is due to the application of the A-operator (11). Note that A can also be inserted at the matrix level, collapsing the presupposition and assertion [SRHF12, Fox13].

(11) $[\![A]\!](\mathrm{p}) = \begin{cases} 1 \text{ iff } p = 1 \\ 0 \text{ iff } p = \# \vee p = 0 \end{cases}$ [Boc81]

Let's consider the unembedded case, the antecedent of conditionals and the scope of non-monotonic quantifiers as case-studies. In each of these, the parse in bold derives the desired

truth-conditions [2]. In the unembedded case (12), obligatory local accommodation (A) at the matrix level (Parse 2) explains the differences between how homogeneity violations and presupposition violations are perceived. In the antecedent of conditional (13), obligatory local accommodation in the antecedent (Parse 2) derives the lack of projection. Finally, in the scope of non-monotonic quantifiers (14), preference for matrix A (Parse 1) over embedded A (Parse 2) explains the fact that homogeneity does project [KC15]. The challenge is how to constrain the distribution of the A-operator such that the correct parses are predicted.

(12)

Parse 1: Mary read the books.	**Parse 2**: A(Mary read the books)
1 iff Mary read all of the books.	1 iff Mary read all of the books.
0 iff Mary read none of the books.	0 iff Mary didn't read all of the books.
# iff Mary read only some.	# never

(13) a. Parse 1: A(If Mary read the books she passed the exam).
1 iff Mary read none or all of the books and if she read all of them she passed.
b. **Parse 2**: If A(Mary read the books), she passed the exam.
1 iff if Mary read all of the books she passed.

(14) a. **Parse 1**: A(Exactly one student read the books).
1 iff Exactly one student read all of the books and all the other students read none.
b. Parse 2: Exactly one student 1 A(t_1 read the books).
1 iff Exactly one student read all of the books.

Enguehard and Chemla propose that there is a general preference for connected propositions and apply this to the distribution of exh [EC21]. They define connectedness as shown in (15), where a proposition is non-connected if there is a world where it is false ordered between worlds where it is true. They propose that parses where the application of exh results in non-connected meanings are dispreferred, which helps explain certain facts about the distribution of exh.

(15) Given an ordering $\leq$ a proposition p is non-connected iff:
$\exists w_1, w_2, w_3$ s.t. $w_1 \leq w_2 \leq w_3 \wedge p(w_1) = 1 \wedge p(w_2) = 0 \wedge p(w_3) = 1$

I propose that connectedness constrains the distribution of the A-operator in a similar way. Given different parses of a sentence with or without the A-operator, I propose that parses which denote a non-connected proposition are dispreferred. Moreover, all else being equal, global accommodation is preferred to inserting A [Hei83] and A is preferred to occur as high as possible. Crucially, connectedness considerations override the preference for global accommodation and for matrix A over embedded A, such that for example if the parse without A is non-connected while the parse with A is connected, the latter is preferred.

In order to apply the notion of connectedness to trivalent propositions, I provide a generalized definition of connectedness in (16). Roughly speaking, a trivalent proposition is non-connected if either the assertive component (a) or the presupposition (b) is non-connected.

(16) **Connectedness for trivalent propositions**: A trivalent proposition p is non-connected iff
a. $\exists w_1, w_2, w_3$ s.t. $w_1 < w_2 < w_3 \wedge p(w_1) = 1 \wedge p(w_2) = 0 \wedge p(w_3) = 1$ **or**
b. $\exists w_1, w_2, w_3$ s.t. $w_1 < w_2 < w_3 \wedge p(w_1) = 1/0 \wedge p(w_2) = \# \wedge p(w_3) = 1/0$

We can now apply this to the problem of homogeneity projection. Consider a context with two books, 1 and 2. Assuming that worlds where Mary read only one of the books are ordered

[2] I assume that presuppositions project universally from the scope of non-monotonic quantifiers

between worlds where she read neither and worlds where she read both, we predict that the parse 1 in (12) is non-connected: there are worlds where p=# ordered between worlds where p=1 and worlds where p=0 (Figure 1, where w_{xy} corresponds to a world where Mary read x and y.). On the other hand, Parse 2 with the A-operator at the matrix level is connected (figure 2), so we successfully predict that the latter is the only available parse in (12). This predicts the fact that homogeneity does not impose any requirements on the common ground.

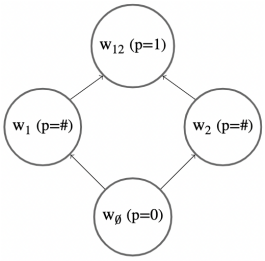


Figure 1: p= Mary read the books

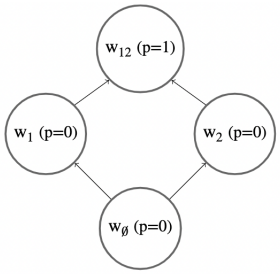


Figure 2: p= A(Mary read the books)

One question that immediately arises is how the ordering of worlds is determined. Given a set of propositions, an ordering of worlds can be generated as illustrated in (17), following [Kra81]. I assume that the ordering in Figure 1 and 2 is determined by the domain alternatives to the sentence with the definite plural. The resulting ordering source for (12) is given in (18).

(17) Given a set of propositions OS, $w_1 <_{\text{OS}} w_2$ **iff**
$\{p \in \text{OS} : p(w_1) = 1\} \subset \{p \in \text{OS} : p(w_2) = 1\}$

(18) OS= { Mary read book 1, Mary read book 2 }

This proposal is similar in spirit to [Fox18], who argues that plausibility consideration mediate the distribution of the A-operator with homogeneity. One crucial difference is that Fox's account predicts that homogeneity projects differently in different contexts, while the connectedness proposal does not predict any context-dependence. In forthcoming work, I provide evidence that homogneity projection does not in fact depend on plausibility.

Turning to the conditional antecedent case, given the same ordering source in (18), the only parse that results in a connected meaning is the one where A is inserted in the antecedent (13-b). This predicts that (13-b) is the only available parse, and therefore that homogeneity doesn't project from the antecedents of conditionals. The parse with matrix A in (13-a) is non-connected (Figure 3): the resulting proposition is false in worlds where *Mary read only some of the books and passed* which are ordered between worlds where *Mary read all of the books and passed* (p=1) and worlds where *Mary read none of the books and passed* (p=1).

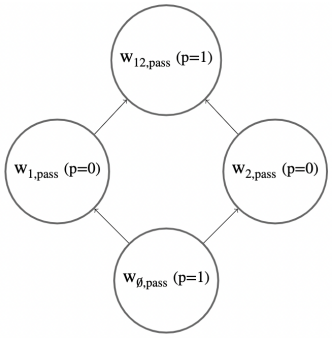


Figure 3: p= A(If Mary read the books, she passed.)

On the other hand, the parse with A in the antecedent (13-b) is connected given the OS in (18), because the worlds where the resulting proposition is false are worlds where Mary read both books but failed, in which both propositions in the ordering source are true. The worlds where p is false are therefore not ordered between any two worlds, making (13-b) connected. Note that the alternative ordering source in (19) would make the parse in (13-b) also non-connected, which would incorrectly predict that global accommodation is preferred, given the general preference for global accommodation. The success of this account therefore depends on the details of the ordering, which have to be motivated more thoroughly in future work.

(19) Alternative OS= { Mary read book 1, Mary read book 2, Mary passed the exam }

Finally, consider a case where homogeneity projects from an embedded environment: the scope of non-monotonic quantifiers. Here, the parse where A is inserted at the matrix level is connected, assuming the ordering source in (20). Given a toy example where there are only two students A and B, the resulting ordering is given in Figure 4, where there is no world that is ordered between the worlds where p=1 (in gray). Given that all else being equal matrix A is preferred over embedded A, we successfully predict that homogeneity does project here.

(20) Ordering source= $\{x \text{ read y } | x \leq_{AT} [\![\text{the students}]\!] \wedge y \leq_{AT} [\![\text{the books}]\!]\}$

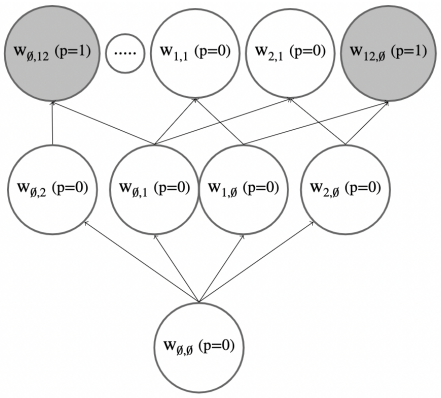


Figure 4: p= A(Exactly one student read the books) (subscript m,n indicates that student A read book m and student B read book n.)

4 Conclusion

After presenting evidence that homogeneity is sensitive to constraints on presupposition accommodation, I proposed that the differences in projection between homogeneity and standard presuppositions can be reduced to a difference in the properties of the proposition that projection gives rise to. Namely, homogeneity projection often results in non-connected propositions which can be made connected by local accommodation. On the other hand, standard presuppositions are generally connected and therefore, unlike homogeneity, do not exhibit this kind of obligatory local accommodation.

The idea behind this proposal is that the disjunctive nature of the homogeneity presupposition results in a type of oddness (formalized as non-connectedness) which can be rescued by inserting A. More broadly, we can take this as a case-study showing that projection can sometimes be misleading as a diagnostic of presuppositions. Constraints on accommodation like PAI offer an alternative diagnostic which does not rely on projection.

References

[AFH22] Athulya Aravind, Danny Fox, and Martin Hackl. Principles of presupposition in development. *Linguistics and Philosophy*, pages 1–42, 2022.

[Boc81] Dmitri A Bochvar. On a three-valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus. *History and Philosophy of Logic*, 2(1-2):87–112, 1981.

[DW22] Omri Doron and Jad Wehbe. A constraint on presupposition accommodation. In *Proceedings of the 23nd Amsterdam Colloquium*, volume 23, 2022.

[EC21] Émile Enguehard and Emmanuel Chemla. Connectedness as a constraint on exhaustification. *Linguistics and Philosophy*, 44(1):79–112, 2021.

[vF04] Kai von Fintel. Would you believe it? the king of france is back! presuppositions and truth-value intuitions. 2004.

[Fox13] Danny Fox. Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. 2013.

[Fox18] Danny Fox. Comments on Križ (2015, 2016). Class notes, 24.979 (MIT, Fox/Heim/Schwarzschild), 2018. http://lingphil.scripts.mit.edu/papers/fox/Kriz_2015_2016.pdf.

[Gaj05] Jon Robert Gajewski. *Neg-raising: Polarity and presupposition*. PhD thesis, Massachusetts Institute of Technology, 2005.

[Hei83] Irene Heim. On the projection problem for presuppositions. *Formal semantics–the essential readings*, pages 249–260, 1983.

[Hei15] Irene Heim. Presupposition. class notes, 24.954 (MIT, Fox/Heim), 2015.

[KC15] Manuel Križ and Emmanuel Chemla. Two methods to find truth-value gaps and their application to the projection problem of homogeneity. *Natural Language Semantics*, 23(3):205–248, 2015.

[Kra81] Angelika Kratzer. The notional category of modality. *Words, worlds, and contexts*, 38:74, 1981.

[Kri15] Manuel Križ. *Aspects of homogeneity in the semantics of natural language*. PhD thesis, University of Vienna, 2015.

[KS21] Manuel Križ and Benjamin Spector. Interpreting plural predication: Homogeneity and non-maximality. *Linguistics and Philosophy*, 44(5):1131–1178, 2021.

[Löb00] Sebastian Löbner. Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy*, 23(3):213–308, 2000.

[Sch93] Roger Schwarzschild. Plurals, presuppositions and the sources of distributivity. *Natural Language Semantics*, 2(3):201–248, 1993.

[Spe13] Benjamin Spector. Homogeneity and plurals: From the strongest meaning hypothesis to supervaluations. Presented at Sinn und Bedeutung 18, 2013.

[SRHF12] Yasutada Sudo, Jacopo Romoli, Martin Hackl, and Danny Fox. Presupposition projection out of quantified sentences: Strengthening, local accommodation and inter-speaker variation. In *Logic, language and meaning*, pages 210–219. Springer, 2012.

Mandarin-speaking children's understanding of *you* 'again' with goal-PPs*

Ting Xu[1], William Snyder[2], and Stella Christie[3]

[1] Tsinghua University, Beijing, China
xuting.thu@gmail.com
[2] University of Connecticut, Storrs, Connecticut, U.S.A.
william.snyder@uconn.edu
[3] Tsinghua University, Beijing, China
christie@mail.tsinghua.edu.cn

Abstract

Like the adverb 'again' in many other languages, Mandarin *you* displays the repetitive/restitutive ambiguity when it modifies a goal-PP (e.g., 'walk to the village'). This paper presents an experimental study which examined 4- and 5-year-old Mandarin-learning children's comprehension of *you* with goal-PPs. The results show that children performed well on the repetitive and restitutive readings, supporting a structural account that would predict no delay in the acquisition of the restitutive reading.

1 Introduction

In many languages, the adverb 'again' exhibits what is called the repetitive/restitutive ambiguity when it modifies a change-of-state predicate, cf. [11, 12, 2]. The difference between these two readings lies in the presupposition: the repetitive reading presupposes a prior event of the same kind, whereas the restitutive reading only presupposes that the result state of the event held before. Here we will focus on the case of goal-PP constructions (by which we mean the combination of a manner-of-motion verb with a PP indicating location/path), as illustrated in (1).

(1) John walked to the village again.
   a. *repetitive*: John had walked to the village before.
   b. *restitutive*: John had been at the village before (e.g., he was born at the village).

Prior research found good performance on restitutive *again* with English goal-PPs by age 4-5, despite extreme scarcity of direct evidence [15]: In a sample of more than 100,000 child-directed utterances, unambiguously restitutive uses of *again* with goal-PPs were entirely absent. The authors proposed English-learning children deduce the restitutive reading from the syntax of English goal-PPs and a basic, repetitive meaning for *again*.

In this study we investigate Mandarin-learning preschoolers' understanding of *you* 'again' with goal-PPs. Like English *again*, *you* permits both repetitive and restitutive readings (2).

(2) Zhangsan you zou dao le na-ge cunzi.
Zhangsan again walk arrive Asp that-CL village
'Zhangsan walked to the village again.' (✓repetitive, ✓restitutive)

*We wish to thank Peng Zhou, Yue Ji, Yue Hu, Honghong Bai, Wenjia Tan, and the reviewers of AC for valuable suggestions and comments. We would also like to thank Anlan Wang, Xueyan Liu and Yijie Zheng for their help with the experimental materials. This work was supported by the National Social Science Fund of China (22BYY076).

Yet, *you* also differs from *again*: (i) *you* always precedes the predicate; and (ii) alongside repetition and restitution it permits temporal continuation (3a), addition (3b), and rhetorical readings (3c).

(3) a. Ta xi-wan yifu you qu zuo fan.
He wash-finish clothes YOU go cook meal
'He did the laundry, and then cooked meals.'
b. Ta congming you qinfen.
He clever YOU hard-working
'He's clever and hard-working.'
c. Ta you bu shi laohu, bu yong pa ta.
He YOU not be tiger not need afraid him
'He's not a tiger. You need not be afraid of him.'

Given that *you* is polysemous and precedes the predicate, it is less clear how its restitutive interpretation is derived and how it is acquired, as will be discussed in Section 2. Acquisitional evidence may provide a way to judge whether *you* is comparable to English *again*. To assess whether Mandarin-learning children understand restitutive readings with goal-PPs by age 4-5, as was found for English, we conducted a comprehension study and found no delay. We discuss the implications for possible sources of *you*'s repetitive/restitutive ambiguity.

2 Background and motivation

Many researchers argue that the repetitive/restitutive ambiguity of 'again' is structural [11, 12, 2]: There is only one 'again', which triggers the presupposition indicated in (4). The repetitive/restitutive readings differ in where *again* adjoins in syntax, as illustrated in (5).[1] Such a structural analysis correctly predicts that when English *again* is preverbal, there is only a repetitive reading (6).

(4) Let P be a property of eventualities and let e be an eventuality.
$[[again]](P)(e)$ is defined only if $\exists e'[P(e') = 1 \& e' < e]$.
Where defined, $[[again]](P)(e) = 1$ iff $P(e) = 1$. (adapted from [12])

(5) a. John walked to the village again.
b. [[John$_1$ [walked [PRO$_1$ to the village]]] again] repetitive
c. [John$_1$ [walked [[PRO$_1$ to the village] again]]] restitutive

(6) John again walked to the village. (✓repetitive, *restitutive)

In Mandarin, since *you* always precedes the predicate, the structural analysis predicts that it should only have a repetitive reading. This prediction is not borne out, as the restitutive interpretation is also allowed (2). Yet, [8, 9, 13, 14] propose that a structural analysis can be maintained: *you* can be generated low, move to a pre-verbal surface position, and later reconstruct. In other words, (2) may be interpreted with multiple scope choices (7): It receives a repetitive reading when it is base-adjoined to the whole proposition (7a), and a restitutive reading when it is base-adjoined to a result-denoting constituent (7b).

(7) a. [again [Zhangsan$_1$ [walk [PRO$_1$ to that village]]]] repetitive
b. [again$_2$ [Zhangsan$_1$ [walk [t$_2$ [PRO$_1$ to that village]]]]] restitutive

[1] [3],[1] treat *to* as synonymous with *at*.

This analysis means learners with knowledge of the repetitive reading will get the restitutive reading for free, as soon as they know both (i) the syntax of goal-PPs and (ii) the syntax of movement and reconstruction for *you*. If children acquire these key characteristics (the nature of goal-PPs and pre-verbal adverbs) prior to repetitive '*you*', we expect simultaneous mastery of repetitive and restitutive *you*. In contrast, an account attributing *you*'s restitutive reading to the existence of a specifically restitutive denotation (à la [6, 10]) would require the learner to rely on direct evidence: uses of *you* in unambiguously restitutive contexts. Given that repetitive *you* undoubtedly occurs more frequently than restitutive *you* in child-directed Mandarin, repetitive *you* should be acquired prior to restitutive *you*. This leads us to ask: Do children understand restitutive *you* as well as they do repetitive *you*?

3 Experiment

3.1 Participants

Participants included 65 Mandarin-learning children from age 3;05 to 5;11 (12 3-year-olds, 38 4-year-olds and 15 5-year-olds, mean age: 4;07) in Beijing, China, who were brought to the lab by their parents and tested individually. We also tested 16 native-speaker adult controls.

3.2 Method and materials

Participants watched a series of stories alongside a puppet who sometimes did not pay attention. The stories were told by the experimenter through cartoon pictures presented in PowerPoint on a laptop. At the end of each story, a puppet popped up on the computer screen and uttered a sentence as a description of what happened in the story, and the participant was asked whether the puppet 'got it right'. If a child participant rejected the test sentence, he/she was asked a follow-up question like, "Why do you think the puppet didn't get it right"? (3) is a sample restitutive item.

(8) *Experimenter*: This is a story about a baby lizard and a baby snail. The lizard hatches under a rock, and the snail is born by a river. The lizard stays under the rock for a while. Then he starts to feel thirsty. So he crawls away from the rock to the river, and gets some water. There he meets the snail. Soon the sun comes out. Feeling hot, the lizard wants to go back under the rock to enjoy the cool shade. He asks his new friend to join him. The snail is sleepy, but decides to follow the lizard and crawls under the rock. Both of them like the cool shade very much, and decide to stay under the rock for a long rest.



First slide of the story



Last slide of the story

Puppet: I know what happened in the story...

a. Xiyi you pa-dao le shitou-xia. (Match)
lizard again crawl-arrive Asp stone-under
The lizard crawled under the rock again.

b. Woniu you pa-dao le shitou-xia. (Mismatch)
snail again crawl-arrive Asp stone-under
The snail crawled under the rock again.

In the story above, the lizard was born under the rock, and thus satisfied the presupposition of restitutive *you* (i.e., the lizard was under the rock before). The other character, the snail, was born near a river and thus did not satisfy the presupposition. There are two possible test sentences for each story: a match sentence such as (8a) and a mismatch one such as (8b). Each participant heard only one test sentence for each story.

We used a 2×2 design: For the story, we manipulated the CONTEXT (2 levels: repetitive and restitutive); for puppet's utterance, we manipulated the TARGET RESPONSE (2 levels: match and mismatch). Participants received 4 trials under each condition. They also received 8 plain goal-PP items, 2 training items and 8 filler items, for a total of 34 items. The goal-PP items and 'again'-items involved the following 8 combinations of manner-of-motion verb and location/path P: *zou-jin* 'walk into', *fei-jin* 'fly into', *tiao-jin* 'jump into', *pao-jin* 'run into', *you-dao* 'swim to', *hua-dao* 'slide to', *piao-dao* 'float to', *pa-dao* 'crawl under'.

4 Results

Of the 65 child participants, 7 children were first excluded from data analysis for low performance (< 75% accuracy) on goal-PP items (n=1) or filler items (n=6). We then excluded data from 23 children who constantly accepted all the *you*-sentences (acceptance rate ≥ 14/16), as if they were ignoring *you*. Finally, we excluded data from one child who constantly rejected the test sentences (rejection rate ≥ 14/16) based on her wrong explanations for why the puppet was wrong (e.g., 'because the lizard wasn't the only one who crawled under the rock'). The explanations suggested the child was both ignoring *you* and calculating an ad-hoc implicature of uniqueness. Our data analysis focused on the remaining 34 children (5 3-year-olds, 22 4-year-olds, 7 5-year-olds, age range: 3;08-5;08; mean: 4;07).

Adults' and the remaining children's acceptance rates are in Figure 1. As expected, for both the repetitive and restitutive stories, the adults overwhelmingly accepted the match items and rejected the mismatch items. Children's performance was highly similar to the adults'.

We fitted mixed effects logistic regression models to the adults data, the child data, and the combined data, with participants' response (1=accept, 0=reject) as the dependent variable, and the intercepts for participants and items as random effects. For the adult model, fixed factors included CONTEXT (repetitive vs. restitutive), TARGET RESPONSE (match vs. mismatch), and their interaction. The model revealed a significant effect of TARGET RESPONSE ($\beta = -10.34$, $p < .001$), but no effect of CONTEXT ($\beta = -1.68$, $p = 0.128$) and no interaction between TARGET RESPONSE and CONTEXT ($\beta = 1.66$, $p = .231$). We fitted models to the child data and compared models with and without CONTEXT, TARGET RESPONSE, AGE (in years as a continuous variable), and their higher order interactions. Model comparison showed no significant improvement in the model's fit when AGE and its interaction with other fixed factors were included; AGE was therefore excluded from the model. The final child model revealed a significant effect of TARGET RESPONSE ($\beta = -2.93$, $p < .001$), but no significant effect of CONTEXT ($\beta = 0.008$, $p = .983$) and no interaction between TARGET RESPONSE and CONTEXT ($\beta = 0.202$, $p = .695$). These results

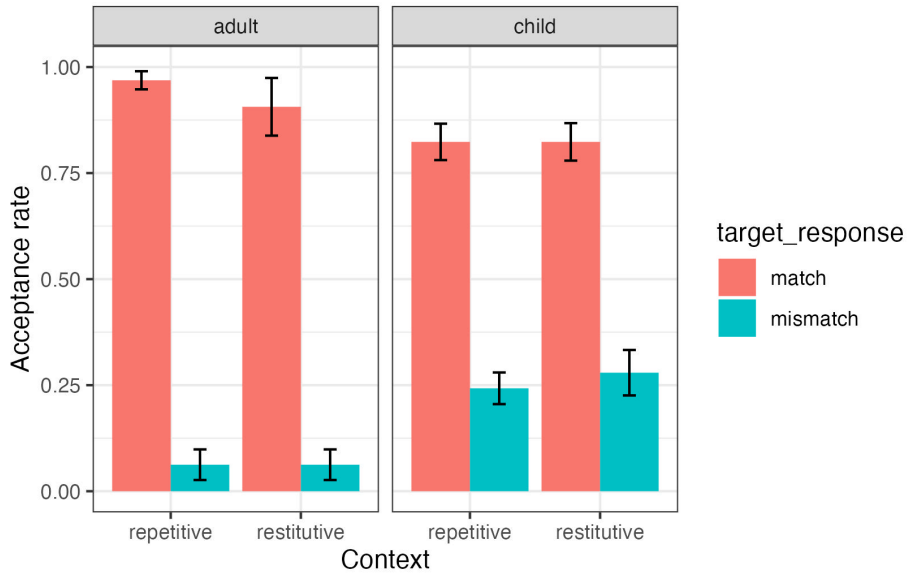


Figure 1: Acceptance rate of adult and child participants. (Error bars indicate standard error.)

suggest that both adults and children distinguished match from mismatch items regardless of the context. For the model fitted on the combined data, fixed factors included CONTEXT, TARGET RESPONSE and PARTICIPANT GROUP (child vs. adult). The model revealed a significant effect of TARGET RESPONSE ($\beta = -2.93$, $p < .001$), PARTICIPANT GROUP ($\beta = -2.12$, $p < .01$), and their interaction ($\beta = 3.89$, $p < .001$). All other p's > .05 NS. The results suggest that both groups made a distinction between match and mismatch items, although adults distinguished match and mismatch items to a greater extent than children.

5 Discussion

5.1 On how Mandarin-learning children acquire restitutive *you*

Our experimental results show that Mandarin-learning 4- and 5-year-olds show good performance on both repetitive and restitutive *you* with goal-PPs. This has implications for developing a theory of how children acquire the restitutive reading of *you*.

In Section 2, we discussed how structural and lexical analyses of the repetitive/restitutive ambiguity of *you* make different predictions for children's acquisition process. A lexical analysis would require the learner to rely on direct evidence: uses of *you* in unambiguously restitutive contexts. Currently we are examining the parental input of two Mandarin-learning children, Tong and Xuexue, whose spontaneous longitudinal data are available in the CHILDES database [16, 5]. While annotation is still in progress, preliminary results suggest that unambiguous restitutive *you* with a goal-PP is not available.[2] The scarcity of direct evidence in child ambient

[2]Our search was limited to goal-PPs, because in certain languages (e.g., French), restitutive 'again' is unavailable with goal-PPs, even though it is available with a lexical accomplishment verb such as 'open'. This means that knowing her language allows restitutive 'again' with a change-of-state predicate does not tell the child that restitutive 'again' is possible with a goal-PP.

speech raises the question of how children can learn the restitutive reading of *you* with goal-PPs if the reading is derived lexically.

Under a structural analysis, on the other hand, learners with knowledge of the repetitive reading will get the restitutive reading for free, as soon as they know both (i) the syntax of the change-of-state predicate and (ii) the characteristic of *you* that allows the adverb to modify a subevent of the complex predicate (i.e., movement and reconstruction for *you*). Given the lack of direct evidence in the input, our experimental finding (i.e., that young preschoolers can access the restitutive reading) favors this analysis.

Notice that the current results are compatible with either simultaneous or sequential acquisition of the two readings (i.e., with the repetitive reading acquired either concurrently with or prior to the restitutive). Data from younger children would help us tease apart these two possibilities. Crucially both possibilities raise the question of what kind of evidence children make use of to master the characteristics of *you*, such as movement and reconstruction. The evidence should be quite general and go beyond goal-PPs or even complex change-of-state predicates. In fact, [8, 9] observed that the ambiguity of *you* is not restricted to change-of-state predicates but can include restructuring verbs, as illustrated in (9).

(9) John you xiangyao guanshang na-shan men.
John again want close that-CL door
'John wants to close that door, and ...'
a. 'he wanted to closed that door before.'
b. 'he close that door before.'
c. 'the door was in a state of being closed before.'

Thus it is possible that the child learns *you* can move and reconstruct based on evidence from restructuring verbs or other change-of-state predicates. We plan to probe this possibility in more detail as our next step.

5.2 On children's indifference to *you*

Despite the success of a number of children, quite a few children behaved as if they were ignoring *you* in our experiment. Why was it easier for children to ignore the presupposition trigger than it was for adults? We speculate that this may be task-related: Our method is adapted from a Truth Value Judgment Task (TVJT, see [4]). Previous studies using the same methodology on children's processing of scalar implicatures (e.g., [7]) have suggested that children are more pragmatically tolerant than adults and thus tend to penalize pragmatic oddity less strictly than adults. As a result, the current task may be less suitable for evaluating children's presuppositions as compared with asserted content. As our next step, we plan to explore other methodologies to better tap into children's knowledge of presupposition.

6 Conclusion

In this study we conducted an experiment to examine Mandarin-learning preschoolers' comprehension of *you* with goal-PPs, and found good comprehension of *you*'s repetitive and restitutive readings. This finding favors a structural analysis of the ambiguity, which would predict no delay in the acquisition of the restitutive reading.

References

[1] Sigrid Beck. There and back again: A semantic analysis. *Journal of semantics*, 22(1):3–51, 2005.

[2] Sigrid Beck and Kyle Johnson. Double objects again. *Linguistic inquiry*, 35(1):97–123, 2004.

[3] Sigrid Beck and William Snyder. The resultative parameter and restitutive again. In Caroline Féry and Wolfgang Sternefeld, editors, *Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow*, Studia grammatica, pages 48–69. Akademie Verlag, Berlin, 2001.

[4] Stephen Crain and Rosalind Thornton. *Investigations in Universal Grammar*. MIT Press, Cambridge, MA, 1998.

[5] Xiangjun Deng and Virginia Yip. A multimedia corpus of child mandarin: The Tong corpus. *Journal of Chinese linguistics*, 46(1):69–92, 2018.

[6] Catherine Fabricius-Hansen. Wi (e) der and again (st). In Caroline Féry and Wolfgang Sternefeld, editors, *Audiatur Vox Sapientiae: A Festschrift for Arnim von Stechow*, Studia grammatica, pages 101–130. Akademie Verlag, Berlin, 2001.

[7] Napoleon Katsos and Dorothy VM Bishop. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition*, 120(1):67–81, 2011.

[8] Yuyang Liu. 'again' skipping in Mandarin Chinese. Unpublished manuscript, 2021.

[9] Yuyang Liu. 'again' skipping in Mandarin Chinese: A syntactic approach. Presentation at the 40th meeting of the West Coast Conference on Formal Linguistics (WCCFL40), 2022.

[10] Walter A Pedersen. A scalar analysis of *again*-ambiguities. *Journal of semantics*, 32(3):373–424, 2015.

[11] Arnim von Stechow. Lexical decomposition in syntax. In Urs Egli, Peter E. Pause, Christoph Schwarze, Arnim von Stechow, and Götz Wienold, editors, *Lexical Knowledge in the Organization of Language*, pages 81–118. John Benjamins, Amsterdam, 1995.

[12] Arnim von Stechow. The different readings of *wieder* 'again': A structural account. *Journal of semantics*, 13(2):87–138, 1996.

[13] Ting Xu. *You* again: how is its ambiguity derived? In Maria Aloni, Vadim Kimmelman, Floris Roelofsen, Galit W. Sassoon, Katrin Schulz, and Matthijs Westera, editors, *Logic, Language and Meaning: 18th Amsterdam Colloquium, Amsterdam, the Netherlands, December 19-21, 2011, Revised Selected Papers (Lecture Notes in Computer Science Volume 7218*, pages 470–479, Amsterdam, 2012. Springer.

[14] Ting Xu. *Almost again: On the semantics and acquisition of decomposition adverbs*. PhD thesis, University of Connecticut, Storrs, CT, 2016.

[15] Ting Xu and William Snyder. There and back again: An acquisition study. *Language acquisition*, 24(1):3–26, 2017.

[16] Li Zhang and Jing Zhou. The development of mean length of utterance in mandarin-speaking children. In Jing Zhou, editor, *The application and development of international corpus-based research methods*, pages 40–58. Education Science Publishing House, 2009.

Epistemic Pseudo-Agreement and Hyper-Domain Semantics for Epistemic Modals

Yichi (Raven) Zhang[1] and Eric Pacuit[2]

[1] Heinrich Heine Universität, Düsseldorf, Germany
yichi.zhang@hhu.de
[2] University of Maryland, College Park, U.S.A.
epacuit@umd.edu

Abstract

We propose a hyper-domain semantics to analyze the group reading of epistemic modals according to which the relevant quantificational base for modals is not represented by a single aggregated information state but rather by a set of information states. The resulting framework allows us to capture what we call "epistemic pseudo-agreements". Additionally, to further investigate the corresponding logic of our hyper-domain semantics, we devise a translation of the consequence relation in our hyper-domain semantics into a new logic called "Logic of Epistemic Disagreement" interpreted over possibility models.

1 Introduction

A key issue for any semantics for epistemic modals [6, 14, 11, 13] is to identify the relevant body of information that supplies a quantificational base for the modal operator. Epistemic disagreements are vivid illustrations of this:

(1) John: Alice might be at the party.
Mary: No, that's not true. She's just tested positive for Covid this morning.

According to [12], there are three different readings of the *might*-claim uttered by John. On John's solipsistic reading, John's assertion of the *might*-claim is warranted since its prejacent is compatible with what John knows. On Mary's solipsistic reading, the quantificational base is supplied by what Mary knows, and since she knows that Alice is not at the party, this is, in part, what warrants her rejection of John's *might*-claim. Finally, there is the group reading in which the quantificational base for the *might*-claim is derived from the *distributed knowledge* of John and Mary. The distributed knowledge of John and Mary is what is known after they each share everything that they know (about who is at the party)—i.e., the distributed knowledge is derived from the intersection of the sets of worlds representing John's private information and the set of worlds representing Mary's private knowledge. Since Alice being at the party is not consistent with John and Mary's pooled information, using the distributed knowledge of John and Mary as the quantificational base of the *might*-claim explains Mary's rejection of the *might*-claim and predicts that John will retract his claim.

One problem with using distributed knowledge for the group reading of epistemic modals is that it incorrectly predicts disagreement about *might*-claims in what we call *epistemic pseudo-agreement* dialogues:

(2) (*Context: John knows that Alice is at the party but has no information on Bob; Mary on the other hand knows that either Alice or Bob is at the party but not both.*)
John: Bob might be at the party.
Mary: I agree.

John's knowledge can be represented by two possible worlds, one in which both Alice and Bob are at the party and one in which Alice is at the party but Bob is not. Mary's knowledge can be represented by two possible worlds, one in which Alice is at the party but Bob is not at the

party, and one in which Bob is at the party but Alice is not at the party. The intersection of these two sets of possible worlds consists of a single possible world in which Alice is at the party but Bob is not at the party. Nevertheless, even though Bob being at the party is inconsistent with the distributed knowledge of John and Mary, Mary agrees with John's *might*-claim.

Von Fintel and Gillies [12] proposed the following additional constraint on the group information state to handle the above problem of pseudo-agreement:

> **Defeasible Closure:** If [the hearer] H knows that φ is compatible with what x knows, for each $x \in G$, then it is reasonable for H to defeasibly infer that φ is compatible with what [the group] G knows.

Assuming that Mary knows what information John has about who is at the party, she can defeasibly infer that Bob being at the party is compatible with what the group knows. So, Defeasible Closure does address the problems that arise from naively using distributed knowledge for group readings of epistemic modals. However, since the fix offered by defeasible closure requires the discourse participants to deliberately draw the inference that φ is compatible with what other people know, its application is limited. For instance, suppose that in (2), Mary has good reasons to suspect that John may not truthfully report what he really knows. In this case, she will not defeasibly infer that Bob's being at the party is compatible with what John knows, even if, in this particular instance, John is indeed being truthful and his information is compatible with Bob being at the party. Nonetheless, it seems she can still assent to John's *might*-claim in (2).

One response to the above discussion is to simply dispense with the group reading of epistemic modals. However, we argue that there are utterances that only make sense with a group reading of epistemic modals. Consider the following sentence adapted from DeRose [2]:

(3) I don't know whether John might have cancer; only the doctor knows.

This sentence is perfectly natural in a context where the doctor has conducted a screening test on John that will conclusively rule out cancer if the test is negative, but will not conclusively affirm that he has cancer if the test is positive. However, this sentence would not make sense if the relevant body of information for the *might*-claim is based on the speaker's knowledge. One plausible interpretation of (3) is that the *might*-claim is based on the doctor's information as illustrated by the following paraphrase:

(4) I don't know whether John's having cancer is compatible with what the doctor knows; only the doctor knows.

The problem with this reading, however, is that it predicts that (5) and (6) have the same meaning, and thus incorrectly predicts that (5) is felicitous.

(5) # I know John doesn't have diabetes, but I don't know whether John might have cancer and diabetes.

(6) I know John doesn't have diabetes, but I don't know whether John's having cancer and diabetes is compatible with what the doctor knows.

Correctly predicting that (5) is infelicitous seems to require a group reading for the *might*-claim taking into account both the doctor's and the speaker's information. Thus, a group reading of the epistemic modals is needed for (4) to be a correct paraphrase of (3) and for (5) to be infelicitous. In this paper, rather than using a single information state to represent group knowledge (e.g., the distributed knowledge of the group), we propose to represent the group information state as a set of information states each corresponding to the private information of a member of the group.

2 Preliminary: Domain Semantics for Epistemic Modals

Our formal model builds on Yalcin's domain semantics [14]. One of the main motivations of domain semantics is to capture epistemic contradictions:

(7) # John doesn't have cancer but he might have cancer.

(8) # Mary supposes that John doesn't have cancer but he might have cancer.

Sentences of the form $\neg p \wedge \Diamond p$ and $p \wedge \Diamond \neg p$ (as well as the variants that swap the order of the conjuncts) intuitively feel contradictory even in an embedded environment (as in (8)). Domain semantics predicts that both (7) and (8) are contradictory. Formulas are interpreted at pairs $\langle w, s\rangle$ where w is a possible world and s is an information state (a set of possible worlds). Given a world-information-state pair $\langle w, s\rangle$, the atomic propositions and the Boolean connectives are interpreted using the possible world w and a valuation function as usual, and modalities quantify over the worlds in the information state s: for instance, $w, s \Vdash \Diamond p$ if, and only if, there exists $w' \in s$ such that $w', s \Vdash p$. That is, $\Diamond p$ tests that the prejacent is compatible with the given body of information specified by s. Say that an information state s *accepts* (or *supports*) a formula φ, denoted $s \vDash \varphi$, if, and only if, $w, s \Vdash \varphi$ for all $w \in s$. Then, sentences such as (7) are unassertable since any information state that accepts $\neg p$ will fail to accept $\Diamond p$, and vice versa.

To see why sentences such as (8) are unassertable, note that the attitude verb 'i suppose' shifts the information state to an information state representing what the agent i supposes: $w, s \Vdash i$ supposes φ if, and only if, $\forall w' \in S_i^w : w', S_i^w \Vdash \varphi$, where S_i^w is the body of information that corresponds to what i supposes at w. It is not difficult to see that this will predict that sentences such as (8) are unassertable. However, as Yalcin noted [14], the semantics sketched above has difficulty explaining why sentences such as (3) of the form 'I don't know whether might p' are *not* contradictory. Furthermore, as pointed by Dorr and Hawthorne [3], shifting to the knowledge state of someone other than the speaker does not help since for all information states s, one of the following must be true: (i) for all $w' \in s$, $w', s \Vdash \Diamond \varphi$ or (ii) for all $w' \in s$, $w', s \Vdash \neg \Diamond \varphi$. Hence, the speaker cannot both fail to know that might p and know that not might p. We address this problem with domain semantics by moving to hyper-domains (i.e., sets of information states) as explained in the next section.

3 Hyper-Domain Semantics for Epistemic Modals

Suppose that $W \neq \varnothing$ and $w \in W$ and $\Sigma \subseteq \wp(W)$. Formulas are interpreted at pairs $\langle w, \Sigma\rangle$, where w is a possible world and Σ is a set of information states (called a hyper-domain). Each element of Σ represents the information possessed by a relevant party in a discourse. We assume that for all $\langle w, \Sigma\rangle$, we have that $w \in \bigcap \Sigma$, so that all discourse participants consider the actual world possible. For an atomic formula p, we have that $w, \Sigma \Vdash p$ iff p is true at w (according to some fixed valuation function mapping atomic propositions to sets of worlds). Compound formulas are evaluated as follows:

$w, \Sigma \Vdash \neg\varphi$ iff $\forall s \in \Sigma : w, \{s\} \nVdash \varphi$; $\quad w, \Sigma \Vdash \varphi \to \psi$ iff $\forall s \in \Sigma : w, \{s\} \nVdash \varphi$ or $w, \{s\} \Vdash \psi$;
$w, \Sigma \Vdash \varphi \wedge \psi$ iff $\forall s \in \Sigma : w, \{s\} \Vdash \varphi$ & $w, \{s\} \Vdash \psi$; $\quad w, \Sigma \Vdash \Diamond\varphi$ iff $\forall s \in \Sigma\, \exists w' \in s$: $w', \{s\} \Vdash \varphi$;
$w, \Sigma \Vdash \varphi \vee \psi$ iff $\forall s \in \Sigma : w, \{s\} \Vdash \varphi$ or $w, \{s\} \Vdash \psi$; $\quad w, \Sigma \Vdash \Box\varphi$ iff $\forall s \in \Sigma\, \forall w' \in s$: $w', \{s\} \Vdash \varphi$.

It is easy to see that the usual inter-definability between the Boolean connectives and the duality between $\Diamond$ and $\Box$ are satisfied. Note that a *might*-claim ($\Diamond\varphi$) is satisfied when every discourse participant deems the prejacent possible, and that the negation of a *might*-claim is satisfied when every discourse participant deems its prejacent impossible. The analogue of Yalcin's notion of *acceptance* is defined as follows: Σ accepts φ, denoted $\Sigma \vDash \varphi$, iff $w, \Sigma \Vdash \varphi$ for all $w \in \bigcup \Sigma$. Now, there is an epistemic disagreement in a hyperdomain Σ when $\Sigma \nVdash \Diamond\varphi$ and $\Sigma \nVdash \neg \Diamond \varphi$. We thus capture the epistemic disagreement in (1) since the prejacent is compatible with John's private

information s_j but not with Mary's s_m, and so the the hyper-dommain $\{s_j, s_m\}$ accepts neither the *might*-claim nor its negation. We also correctly predict the epistemic pseudo-agreement in (2): since both John's and Mary's private information states contain a world where Bob is at the party, the *might*-claim is accepted. Finally, semantic consequence, denoted $\models^{hd}$, is defined as preservation of acceptance at all hyper-domains: $\Gamma \models^{hd} \varphi$ iff for all W, $V : \mathsf{At} \to \wp(W)$ and $\Sigma \subseteq \wp(W)$, if $\Sigma \models \gamma$ for all $\gamma \in \Gamma$, then $\Sigma \models \varphi$. Hence, we have that $\neg p \wedge \Diamond p \models^{hd} \bot$ as every hyper-domain that accepts $\neg p$ will fail to accept $\Diamond p$ (similarly for the other variants of epistemic contradictions). In addition, as with domain semantics, our hyper-domain semantics collapses strings of modals to the innermost modality: for instance, we have that $\Diamond \Box \Diamond p \models^{hd} \Diamond p$.

4 Logic of Epistemic Disagreement (LED)

To investigate the logic of our hyper-domain semantics, we take inspiration from Schulz's [9] translation of the consequence relation of domain semantics into the modal logic **S5**. The target logical system for our translation of the semantic consequence for hyper-domain semantics is a bimodal logic interpreted in models based on *partial states*, also called *possibilities* or *situations*, where the truth value of a formula may be unsettled [5, 10, 4]. The two modalities included in our language are the usual epistemic modality $\Box$ representing "must" and a "consensus" modality $[C]$ indicating that the group collectively agrees on the prejacent. For simplicity, we assume that $[C]$ is always the outermost modality and does not iterate—i.e., it does not embed under either $[C]$ or $\Box$. More formally, the language $\mathcal{L}_{LED}$ is defined recursively as follows: $\varphi ::= \psi \mid [C]\psi \mid -\varphi \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \supset \varphi$, where $\psi ::= p \mid -\psi \mid \psi \wedge \psi \mid \psi \vee \psi \mid \psi \supset \psi \mid \Box\psi$ for $p \in \mathit{Atoms}$ (the set of atomic propositions). We define $\Diamond\psi$ as $-\Box-\psi$ and $\langle C\rangle\psi$ as $-[C]-\psi$.

Formulas of $\mathcal{L}_{LED}$ are interpreted in a possibility model $\mathcal{M} = \langle S, \sqsubseteq, C, V\rangle$ where S is a non-empty set, $\sqsubseteq$ is a partial order on S (reflexive, transitive and anti-symmetric relation on S), C is a binary relation on S, and V is a valuation that assigning to ever atomic formula a nonempty subset U of S satisfying persistence and refinability [4]:

Persistence: if $x \in U$ and $y \sqsubseteq x$, then $y \in U$;

Refinability: if $x \notin U$, then $\exists y \sqsubseteq x\ \forall z \sqsubseteq y : z \notin U$.

Intuitively, y is a refinement of x ($y \sqsubseteq x$) when y settles all the the formulas that x does and possibly more. Persistence guarantees that every atomic proposition that is settled true/false in x is also settled true/false in y. Refinability guarantees that if an atomic proposition not settled true at x, then there is a refinement that settles it false. The intended interpretation of $x\ C\ y$ is that y is a relevant body of information that corresponds to some discourse participant's private information in the *context* represented by x. We impose the following constraint on $\sqsubseteq$ and C mirroring the constraint on hyper-domains requiring that every information state in the hyper-domain contains the actual world: for all $x, y \in S$, if $x\ C\ y$, then $x \sqsubseteq y$. That is, if x represents the context of the conversation, then every participant's information at x must be able to be refinable into x.

Truth of a formula $\varphi \in \mathcal{L}_{LED}$ at a state x in a model $\mathcal{M}$ is defined as follows: For any atomic formula p, $\mathcal{M}, x \Vdash p$ iff $x \in V(p)$.

$\mathcal{M}, x \Vdash -\varphi$ iff $x \nVdash \varphi$

$\mathcal{M}, x \Vdash \varphi \wedge \psi$ iff $\mathcal{M}, x \Vdash \varphi$ and $\mathcal{M}, x \Vdash \psi$

$\mathcal{M}, x \Vdash \varphi \vee \psi$ iff $\mathcal{M}, x \Vdash \varphi$ or $\mathcal{M}, x \Vdash \psi$

$\mathcal{M}, x \Vdash \varphi \supset \psi$ iff either $\mathcal{M}, x \nVdash \varphi$ or $\mathcal{M}, x \Vdash \psi$

$\mathcal{M}, x \Vdash \Box\varphi$ iff $\forall y \sqsubseteq x$: $\mathcal{M}, y \Vdash \varphi$

$\mathcal{M}, x \Vdash [C]\varphi$ iff $\forall y$: if $x\ C\ y$ then $\mathcal{M}, y \Vdash \varphi$

Note that, unlike in standard possibility semantics where persistence holds for all formulas, $\Diamond$-formulas may fail to be persistent: a *might*-claim may hold at a situation but fail to hold at one of its refinements. The definitions of validity on a frame $\langle S, \sqsubseteq, C\rangle$ and a semantic consequence $\models^{led}$ are defined as usual.

We provide a translation scheme for a portion of the language of hyper-domain semantics in which (i) there are no iterated modalities (i.e., no formulas of the form $\Box\varphi$ where φ contains a modality) and (ii) there are no modalities in the scope of a disjunction or implication (e.g., no formula of the form $\Diamond\varphi \vee \psi$ or $\Box\varphi \supset \psi$). Suppose that φ is such a formula of hyper-domain semantics, the formula $\tau(\varphi) \in \mathcal{L}_{LED}$ is defined in two steps: (1) transform φ into a logically equivalent formula φ' where all negations scope below any modal; and (2) define $\tau'(\varphi')$ recursively: (i) $\tau'(p) = p$; (ii) $\tau'(\neg\varphi') = \Box{-}\tau'(\varphi')$; (iii) $\tau'(\varphi' \wedge \psi') = \tau'(\varphi') \wedge \tau'(\psi')$; (iv) $\tau'(\varphi' \vee \psi') = \Box\Diamond(\tau'(\varphi') \curlyvee \tau'(\psi'))$; (v) $\tau'(\varphi' \to \psi') = \Box(\tau'(\varphi') \supset \tau'(\psi'))$; (vi) $\tau'(\Box\varphi') = \Box\tau'(\varphi')$; and (vii) $\tau'(\Diamond\varphi') = \Diamond\tau'(\varphi')$. For example, $\tau(\neg\Box p) = \tau'(\Diamond\neg p) = \Diamond\Box{-}p$. Our first main result is τ preserves validity: for all formulas φ and sets of formulas Γ (in the restricted language), $\Gamma \models^{hd} \varphi$ iff $[C]\tau(\Gamma) \models^{led} [C]\tau(\varphi)$, where $[C]\tau(\Gamma)$ means $[C]\tau(\gamma)$ for all $\gamma \in \Gamma$ (see the Appendix).

The axiomatization of LED for language $\mathcal{L}_{LED}$ includes the following: (i) all axiom schemas from classical logic; (i) the **S4** axioms for $\Box$; (iii) the axiom scheme corresponding to the refinability $\Box\Diamond\varphi \supset \varphi$; (iv) a limited version of persistence $\alpha \supset \Box\alpha$, where α is a *positive formula*; (vi) the **KD** axioms for $[C]$; and (vii) $\varphi \supset [C]\Diamond\varphi$. The last axiom, called *entertainability*, says that if φ is true, then everyone should agree that φ is at least possible. This axiom corresponds to the aforementioned condition on C stating that if $x\ C\ y$, then $x \sqsubseteq y$. Importantly, the T axiom is not satisfied for $[C]$, in particular, sentences of the form $[C]\Diamond p \supset \Diamond p$ may not be true at a partial state. This is what happens in pseudo-agreement situations, such as (2), in which it is possible for everybody at a situation x to agree that $\Diamond p$ while p is actually false at x, and so is $\Diamond p$ is false at x. The logic is closed under Modus Ponens and necessitation for both $\Box$ and $[C]$. Some details of the proof of completeness can be found in the Appendix.

We conclude this section by discussing the consequences of imposing additional constraints concerning the interaction of C and $\sqsubseteq$. A full discussion of these constraints is left for the full paper. Given the restriction on the language that $[C]$ cannot be embed under $\Box$, the formulas corresponding to the constraints discussed below cannot be stated in the language $\mathcal{L}_{LED}$, but they do help to refine the notion of *context* we employ. We start with the following two constraints from [4]:

R-rule: if $x' \sqsubseteq x$, $x'\ C\ y'$, and $y' \between z$, then $\exists y : xCy$ and $y \between z$.
R⇒win: if $x\ C\ y$, then $\forall y' \sqsubseteq y \exists x' \sqsubseteq x \forall x'' \sqsubseteq x' \exists y'' : x''\ C\ y''$ and $y'' \between y'$.

For any two possibilities x and y, $x \between y$ means that x and y have a common refinement. The above two conditions guarantee that sets of partial states associated with $[C]$-formulas satisfy persistence and refinability. Persistence of $[C]$-formulas means that if every participant at a context x agrees on some formula φ (e.g., $\Diamond p$), then at all refinements of x, everybody should still agree on φ. This requires that the group of relevant discourse participants remain fixed across refinements. It also means that a refinement of a context is not accompanied by the discourse participants' gaining new information. Otherwise, there could be some discourse participant's information that is compatible with p, but no longer compatible after the discourse participant acquires new information (e.g., that p is false). Thus, our notion of context refinement is different from the standard notion of context *update*. Our proposal is to view a refinement of a context as an update with new relevant issues or questions under discussion [8, 7]. That is, if y is a refinement of x, then y may settle additional issues that are deemed irrelevant at x. For example, suppose that Alice, Bob, and Carla are at the party. If x represents a context in which the question under discussion only concerns the whereabouts of Alice and Bob, then a refinement of x may represent a context where the question under discussion concerns the whereabouts of all three individuals.

Finally, we consider three additional constraints from [10]. The first two constraints that one might impose are:

C(ontinuous): If $x' \sqsubseteq x$ and $x' \; C \; y'$, then $\exists y : x \; C \; y$ and $y' \sqsubseteq y$.

O(pen): If $x \; C \; y$ and $y' \sqsubseteq y$, then $\exists x' \sqsubseteq x : x' \; C \; y'$.

While it is clear that **C** should be satisfied given our intended interpretation of C and our understanding of context discussed above, imposing both **O** and our **E**(ntertainability) condition (if $y \; C \; x$, then $x \sqsubseteq y$) leads to a counterintuitive constraint on our models. Given **O** and **E**, it follows that if $x \sqsubseteq y$ and $y' \sqsubseteq y$, then $\exists x' : x' \sqsubseteq x$ and $x' \sqsubseteq y'$. That is, every two refinements of a single situation have a common refinement. This means that we cannot impose both persistence and refinability. For this reason, we are tempted to drop the constraint **O**. However, for now we just highlight the tension between **O** and **E** and leave additional investigation to future work.

The third constraint that might be imposed is:

Conv(ergence): If $x' \sqsubseteq x$ and $x \; C \; y$, then $\exists y' \sqsubseteq y : x' \; C \; y'$.

Conv guarantees the persistence of formulas of the form $\langle C \rangle \varphi$ (provided that φ is persistent). This squares with the current interpretation of context refinement: if someone endorses that φ at x then the same person should still endorse φ at a refinement of x' of x. In our ongoing work, we will explore other interactions between C and $\sqsubseteq$ and lift the syntactic restriction currently imposed on $\mathcal{L}_{LED}$ leading to a translation of the full language of hyper-domain semantics.

5 Adding Knowledge Operator

In this section, we define a knowledge operator in hyper-domain semantics. Rather than representing an agent's knowledge as a set of worlds, we use a hyper-domain. Consider (3) again: in order to make "I don't know that John might have cancer" and "I don't know that it's not the case that John might have cancer" both true (i.e., $w, \Sigma \Vdash \neg K_i \Diamond p$ and $w, \Sigma \Vdash \neg K_i \neg \Diamond p$), we assume that the speaker's knowledge $\mathcal{K}_i^w$ contains two information states s_1 and s_2 where s_1 is compatible with John's having cancer whereas s_2 is not. In this case, the speaker internalizes two different epistemic perspectives and her knowledge state is indeterminate between them (cf. [13]). The knowledge operator K_i is defined as follows:

- $w, \Sigma \Vdash K_i \varphi$ iff $\forall w' \in \bigcup \mathcal{K}_i^w : w', \Sigma \cup \mathcal{K}_i^w \Vdash \varphi$, where $\mathcal{K}_i^w$ is the hyper-domain that corresponds to what i knows at w.

Note that K_i shifts the hyper-domain for evaluating its prejacent to the combined state $\Sigma \cup \mathcal{K}_i^w$ rather than agent i's knowledge state. This ensures that the truth axiom $K_i \varphi \supset \varphi$ is accepted at any hyper-domain. In particular, we have that $K_i \Diamond p \supset \Diamond p$ is accepted at any hyper-domain. This is not true in standard domain semantics since p may be compatible with what the knowledge ascribee knows, but incompatible with what the interlocutor knows (see [1] for discussion of a related problem).

Now, we can unpack $w, \Sigma \Vdash \neg K_i \Diamond p$ and $w, \Sigma \Vdash \neg K_i \neg \Diamond p$ into the following conditions:

- $w, \Sigma \Vdash \neg K_i \Diamond p$ iff $\forall s \in \Sigma . \exists s' \in \{s\} \cup \mathcal{K}_i^w . \forall w' \in s' : w', \{s'\} \nVdash p$
- $w, \Sigma \Vdash \neg K_i \neg \Diamond p$ iff $\forall s \in \Sigma . \exists s' \in \{s\} \cup \mathcal{K}_i^w . \exists w' \in s' : w', \{s'\} \Vdash p$

Since $\mathcal{K}_i^w$ contains two information states s_1 and s_2 each of which witnesses one of the two conditions, we correctly predict $w, \Sigma \Vdash \neg K_i \Diamond p$ and $w, \Sigma \Vdash \neg K_i \neg \Diamond p$ to be jointly satisfiable.

6 Conclusion

In this abstract, we presented a hyper-domain semantics that offers a straightforward account of the group reading of epistemic modals. Combining the idea of representing the group information directly using a set of information states with a Yalcin-style domain semantics, we are also able to define a knowledge operator that resolves certain problems associated with the standard domain semantics. We then show that our hyper-domain semantics can be axiomatized via the translation into a new bimodal possibility semantics.

References

[1] Bob Beddor and Simon Goldstein. Mighty knowledge. *The Journal of Philosophy*, 118:229 – 269, 2021.

[2] Keith DeRose. Epistemic possibility. *Philosophical Review*, 100:581 – 605, 1991.

[3] Cian Dorr and John Hawthorne. Embedding epistemic modals. *Mind*, 122:867 – 913, 2013.

[4] Wesley H. Holliday. Possibility semantics. In Melvin Fitting, editor, *Selected Topics from Contemporary Logics*, pages 108–130. College Publications, 2021.

[5] Lloyd Humberstone. From worlds to possibilities. *Journal of Philosophical Logic*, 10:313 – 339, 1981.

[6] Angelika Kratzer. *Modals and Conditionals: New and Revised Perspectives*. Oxford University Press, Oxford, 2012.

[7] Craige Roberts. Information structure in discourse. OSU Working Papers in Linguistics, volume 49, J. Yoon and A. Kathol (eds.), 1996.

[8] Craige Roberts. Information structure: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, 5:1 – 69, 2012.

[9] Moritz Schulz. Epistemic modals and informational consequence. *Synthese*, 174:385 – 395, 2010.

[10] Johan van Benthem, Nick Bezhanishvili, and Wesley H. Holliday. A bimodal perspective on possibility semantics. *Journal of Logic and Computation*, 27:1353– 1389, 2017.

[11] Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:220 – 260, 1996.

[12] Kai von Fintel and Anthony Gillies. 'Might' made right. In Andy Egan and Brian Weatherson, editors, *Epistemic Modality*, pages 108–130. Oxford University Press, 2011.

[13] Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122:45 – 92, 2013.

[14] Seth Yalcin. Epistemic modals. *Mind*, 116:983–1026, 2007.

A Formal Results

Theorem A.1. *The translation scheme τ preserves validity in the following way: $\Gamma \vDash^{hd} \varphi$ iff $[C]\tau(\Gamma) \vDash^{led} [C]\tau(\varphi)$ where $[C]\tau(\Gamma)$ abbreviates $[C]\tau(\gamma)$ for all $\gamma \in \Gamma$.*

Proof. We prove this by first proving **Proposition A.3**: given a pointed hyper-domain model $\langle W, V\rangle, \Sigma$ and its corresponding pointed possibility model $\mathcal{M}, x$, the following holds: $\Sigma \vDash \varphi$ iff $\mathcal{M} = \langle S, C, \sqsubseteq, V'\rangle, x \Vdash [C]\tau(\varphi)$. The correspondence is defined as follows: given a pair $\langle W, V\rangle, \Sigma$, we can generate a pair $\mathcal{M}, x$ by making (i) $x = \bigcap \Sigma$; (ii) $S = \wp(W) - \{\varnothing\}$; (iii) $y' \sqsubseteq y$ iff $y' \subseteq y$; (iv) xCy iff $y \in \Sigma$; (v) $V'(p) = \{s \in S \mid s \subseteq V(p)\}$. We can check that the resulting model is indeed of the right kind; for instance, for all $x, y \in S$, we have that if $x\ C\ y$, then $x \sqsubseteq y$ since $\bigcap \Sigma \subseteq y$ for any $y \in \Sigma$. Borrowing a term from [10], let us call the resulting $\mathcal{M}, x$ the posibilization of $\langle W, V\rangle, \Sigma$. As for the opposite direction going from $\mathcal{M}, x$ to $\langle W, V\rangle, \Sigma$, it suffices to show that $\mathcal{M}, x$ is isomorphic to the possibilization of $\langle W, V\rangle, \Sigma$. Given **Proposition A.3**, **Theorem A.1** immediately follows. Now to prove **Proposition A.3**, we will first prove **Proposition A.2**, the special case where Σ is a singleton hyper-domain. □

Proposition A.2. *Suppose that $W \neq \varnothing$, $V : Atoms \to \wp(W)$ and $s \subseteq W$. Let $\mathcal{M} = \langle S, C, \sqsubseteq, V\rangle$ be a model where (i) $x = \bigcap\{s\} = s$; (ii) $S = \wp(W) \setminus \{\varnothing\}$; (iii) $y' \sqsubseteq y$ iff $y' \subseteq y$; (iv) xCy iff $y \in \{s\}$; (v) $V'(p) = \{s \in S \mid s \subseteq V(p)\}$. Then, $\{s\} \vDash \varphi$ iff $\mathcal{M}, x \Vdash \tau(\varphi)$, where $x = s$.*

Proof. First, since $x = s$, *(iv)* implies that $x\ C\ x$. Let φ' be a formula equivalent to φ but with negations all scoping below modals, and all strings of modals collapsed. Then we can show $\{s\} \vDash \varphi$ iff $\mathcal{M}, x \Vdash \tau'(\varphi')$ by an induction on the structure of φ' by using induction hypotheses: for all $x' \in \wp(W) \setminus \{\varnothing\}$, $\{x'\} \vDash \alpha$ iff $\mathcal{M}, x' \Vdash \tau'(\alpha)$. Consider the left to right direction first.

1. Let φ' be of the form p: then $\tau'(p) = p$. Since $\{s\} \vDash \varphi'$, we have $\forall w \in \bigcup\{s\} : w, \{s\} \Vdash p$, which means $\forall w \in s : w \in V(p)$. Hence, $s \subseteq V(p)$ which means $\mathcal{M}, x \Vdash p$.

2. Let φ' be of the form $\neg\alpha$: then $\tau'(\neg\alpha) = \Box - \tau'(\alpha)$. Since $\{s\} \vDash \neg\alpha$, it follows that $\forall w \in s : w, \{s\} \Vdash \neg\alpha$ and thus $\forall w \in s : w, \{s\} \nVdash \alpha$. Now, assume $\mathcal{M}, x \nVdash \Box - \tau'(\alpha)$ for contradiction. Then, $\exists x' \sqsubseteq x : \mathcal{M}, x' \Vdash \tau'(\alpha)$. Given that $x' \sqsubseteq x$ iff $x' \subseteq x$ and $x = s$, this just amounts to $\exists x' \subseteq s : \mathcal{M}, x' \Vdash \tau'(\alpha)$. By the induction hypothesis, we derive $\exists x' \subseteq s : \{x'\} \vDash \alpha$. Then, $\exists x' \subseteq s \forall w' \in x' : w', \{x'\} \Vdash \alpha$. Now recall that φ' is a formula where all negations scope below modals, this means that α does not contain any modal operators. Given this, $\exists x' \subseteq s \forall w' \in x' : w', \{x'\} \Vdash \alpha$ contradicts $\forall w \in s : w, \{s\} \nVdash \alpha$ since the truth of Boolean formulas do not depend on the hyper-domain.

3. Let φ' be of the form $\alpha \wedge \beta$: then $\tau'(\alpha \wedge \beta) = (\tau'(\alpha) \wedge \tau'(\beta))$. Given $\{s\} \vDash \alpha \wedge \beta$, $\mathcal{M}, x \Vdash (\tau'(\alpha) \wedge \tau'(\beta))$ follows straightforwardly from the induction hypothesis.

4. Let φ' be of the form $\alpha \vee \beta$: then $\tau'(\alpha \vee \beta) = \Box \Diamond (\tau'(\alpha) \curlyvee \tau'(\beta))$. Given $\{s\} \vDash \alpha \vee \beta$, it follows that $\forall w \in s : w, \{s\} \Vdash \alpha \vee \beta$, which means $\forall w \in s : w, \{s\} \Vdash \alpha$ or $w, \{s\} \Vdash \beta$. We also have $\forall w \in s :$ if $w, \{s\} \nVdash \alpha$ then $w, \{s\} \Vdash \beta$, which entails if $\forall w \in s : w, \{s\} \nVdash \alpha$ then $\forall w \in s : w, \{s\} \Vdash \beta$, which amounts to if $\{s\} \vDash \neg\alpha$ then $\{s\} \vDash \beta$. From $\{s\} \vDash \beta$ by induction hypothesis, we have $\mathcal{M}, s \vDash \tau'(\beta)$; from $\{s\} \vDash \alpha$, since $\neg\alpha$ does not contain any modal operators, by what we have just shown in step (2) above, we have $\mathcal{M}, s \Vdash \Box - \tau'(\alpha)$. Hence, we can rewrite the conditional as: if $\mathcal{M}, s \Vdash \Box - \tau'(\alpha)$ then $\mathcal{M}, s \vDash \tau'(\beta)$. First assume $\mathcal{M}, s \Vdash \Box - \tau'(\alpha)$, then we immediately derive $\mathcal{M}, s \Vdash \tau'(\beta)$. Since β does not contain any modals, by persistence, $\forall s' \sqsubseteq s : \mathcal{M}, s' \Vdash \tau'(\beta)$. It then follows that $\mathcal{M}, s \Vdash \Box \Diamond (\tau'(\alpha) \curlyvee \tau'(\beta))$. Now assume $\mathcal{M}, s \nVdash \Box - \tau'(\alpha)$ instead. Then, $\exists s' \sqsubseteq s : \mathcal{M}, s' \Vdash \tau'(\alpha)$, which means $\exists s' \sqsubseteq s : \mathcal{M}, s' \Vdash \tau'(\alpha) \curlyvee \tau'(\beta)$. Since α does not contain modals, by persistence, we have $\mathcal{M}, s \Vdash \Box \Diamond (\tau'(\alpha) \curlyvee \tau'(\beta))$ again.

5. The proof for $\rightarrow$ is analogous to the proof for $\vee$.

6. Let φ' be of the form $\Diamond\alpha$: then $\tau'(\Diamond\alpha) = \Diamond\tau'(\alpha)$. Since $\{s\} \vDash \Diamond\alpha$, it follows that $\forall w \in s : w, \{s\} \Vdash \Diamond\alpha$ which just means $\exists w \in s : w, \{s\} \Vdash \alpha$. Assume $\mathcal{M}, x \nVdash \Diamond\tau'(\alpha)$ for contradiction, from which it follows that $\forall x' \subseteq x : \mathcal{M}, x' \nVdash \tau'(\alpha)$. By the induction hypothesis, we derive $\forall x' \subseteq x : \{x'\} \nvDash \alpha$. Then, $\forall x' \subseteq x \exists w' \in x' : w', \{x'\} \nVdash \alpha$. Since $x = s$, this contradicts $\exists w \in s : w, \{s\} \Vdash \alpha$. To see this, consider the singleton states $\{w'\} \subseteq x$.

7. Let φ' be of the form $\Box\alpha$: then $\tau'(\Box\alpha) = \Box\tau'(\alpha)$. Since $\{s\} \vDash \Box\alpha$, it follows that $\forall w \in s : w, \{s\} \Vdash \alpha$. Assume $\mathcal{M}, x \nVdash \Box\tau'(\alpha)$ for contradiction; it follows that $\exists x' \subseteq x : \mathcal{M}, x' \nVdash \tau'(\alpha)$. By the induction hypothesis, we derive $\exists x' \subseteq x : \{x'\} \nvDash \alpha$. Unpacking gives us $\exists x' \subseteq x \exists w' \in x' : w', \{x'\} \nVdash \alpha$, which $\forall w \in s : w, \{s\} \Vdash \alpha$ given that α is Boolean and the truth of Boolean formulas does not depend on the hyper-domain.

The right to left direction is analogous. □

Proposition A.3. *Suppose that $W \neq \varnothing$, $V : Atoms \to \wp(W)$ and $s \subseteq W$. Let $\mathcal{M} = \langle S, C, \sqsubseteq, V\rangle$ be the possibilization of $\langle W, V\rangle, \Sigma$. Then, $\Sigma \vDash \varphi$ iff $\mathcal{M}, x \Vdash [C]\tau(\varphi)$.*

Proof. We will illustrate with the left to right direction: if $\Sigma \vDash \varphi'$ then $\mathcal{M}, x \Vdash [C]\tau'(\varphi')$.

1. Let φ' be atomic; we want to show $\mathcal{M}, x \Vdash [C]p$. Since $\Sigma \vDash p$, we have $\forall w \in \bigcup\Sigma : w, \Sigma \Vdash p$, which means $\forall w \in \bigcup\Sigma : w \in V(p)$. Assume $\mathcal{M}, x \nVdash [C]p$ for contradiction. It follows that $\exists y : xCy \;\&\; \mathcal{M}, y \nVdash p$, which means $y \notin V(p)$. Since $y \in \Sigma$, this contradicts $\forall w \in \bigcup\Sigma : w \in V(p)$. Hence, $\mathcal{M}, x \Vdash [C]p$.

2. Let φ' be of the form $\neg\alpha$; we want to show $\mathcal{M}, x \Vdash [C]\,\Box - \tau'(\alpha)$. Since $\Sigma \vDash \neg\alpha$, we have $\forall w \in \bigcup\Sigma : w, \Sigma \Vdash \neg\alpha$, which means $\forall w \in \bigcup\Sigma \forall s \in \Sigma : w, \{s\} \nVdash \alpha$. By the definition of acceptance again, it follows that $\forall s \in \Sigma : \{s\} \vDash \neg\alpha$. Given Proposition A.2, we have $\forall s \in \Sigma : \mathcal{M}', s \Vdash [C]\,\Box - \tau'(\alpha)$, where $\mathcal{M}', s$ is the possibilization of $\langle W, V\rangle, \{s\}$. Now since $\{s\}$ only contains one state, it means if $s\ C\ y$ then $y = s$, and given $\mathcal{M}', s \Vdash [C]\,\Box - \tau'(\alpha)$, it follows that $\mathcal{M}', s \Vdash \Box - \tau'(\alpha)$. Since $\mathcal{M}'$ and $\mathcal{M}$ are both generated from $\langle W, V\rangle$, they can only differ possibly with respect to C. But since α does not contain any $[C]$, $\forall s \in \Sigma : \mathcal{M}, s \Vdash \Box - \tau'(\alpha)$. This implies $\mathcal{M}, x \Vdash [C]\,\Box - \tau'(\alpha)$.

3. Proofs for all the other operators can be given along the same lines, for example:

   Let φ' be of the form $\Diamond\alpha$; we want to show $\mathcal{M}, x \Vdash [C]\,\Diamond\,\tau'(\alpha)$. Since $\Sigma \vDash \Diamond\alpha$, we have $\forall w \in \bigcup\Sigma . \forall s \in \Sigma . \exists w' \in s : w', \{s\} \Vdash \alpha$, from which it follows that $\forall s \in \Sigma : \{s\} \vDash \Diamond\alpha$. Given Proposition A.2, we have $\forall s : \mathcal{M}', s \Vdash [C]\,\Diamond\,\tau'(\alpha)$. Since s only C-relates to itself in $\mathcal{M}'$, we have $\forall s : \mathcal{M}', s \Vdash \Diamond\tau'(\alpha)$. Because $\mathcal{M}'$ and $\mathcal{M}$ are again generated from the same $\langle W, V\rangle$, it follows that $\forall s \in \Sigma : \mathcal{M}, s \Vdash \Diamond\tau'(\alpha)$. Hence, $\mathcal{M}, x \Vdash [C]\,\Diamond\,\tau'(\alpha)$.

The right to left direction can be established analogously. □

Definition A.4. The canonical model for LED is defined as $\mathfrak{M}^\Lambda = \{S^\Lambda, \sqsubseteq^\Lambda, C^\Lambda, V^\Lambda\}$, where (i) S^Λ is the set of all consistent and deductively closed sets of formulas of LED; (ii) $\Gamma' \sqsubseteq^\Lambda \Gamma$ iff $\{\varphi \mid \Box\varphi \in \Gamma\} \subseteq \Gamma'$, that is, for all formulas φ, $\Box\varphi \in \Gamma$ implies $\varphi \in \Gamma'$; (iii) $\Gamma' \in C^\Lambda(\Gamma)$ iff $\{\varphi \mid [c]\varphi \in \Gamma\} \subseteq \Gamma'$; (iv) for any atomic formula p, $V^\Lambda(p) = \{\Gamma \in S^\Lambda \mid p \in \Gamma\}$.

Theorem A.5 (Soundness and Strong Completeness)**.** *$\Gamma \vdash^{led} \varphi$ if and only if $\Gamma \vDash^{led} \varphi$.*

Be surprised whether: question-to-cleft reduction*

Ziling Zhu[1] and Yimei Xiang[2]

[1] Rutgers University, New Brunswick, New Jersey, U.S.A.
ziling.zhu@rutgers.edu
[2] Rutgers University, New Brunswick, New Jersey, U.S.A.
yimei.xiang@rutgers.edu

Abstract

We propose a question-to-cleft reduction account to explain why emotive factives (e.g., *be surprised*) cannot embed *whether* questions. First, drawing on the distinct anaphor-licensing patterns of *wh-* and *whether*-questions, we discern discourse referents that may serve as short answers to questions, termed 'answer discourse referents (ans-drefs)'. Next, we argue that questions embedded under emotive factives are always reduced to clefts. This reduction requires to extract a unique ans-dref from the embedded question, which is possible in *wh*-embeddings but not in *whether*-embeddings.

1 Introduction

Cognitive factives (e.g., *know*) and emotive factives (e.g., *be surprised*, *be happy*) have different selectional patterns (Karttunen 1977): as exemplified in (1) and (2), while both types of factives can embed *wh*-questions (*wh*Qs), only cognitive factives can embed *whether*-questions, which can be interpreted either as polar questions (PolQs) or alternative questions (AltQs).

(1) a. Jane knows who danced.
b. Jane knows whether Andy or Billy danced.

(2) a. Jane is surprised who danced.
b. *Jane is surprised whether Andy or Billy danced.

Existing analyses typically attribute this distinction to semantic restrictions of the factive predicates (d'Avis 2002; Abels 2004; Romero 2015) or to a pragmatic competition among forms of questions (Sæbø 2005; Guerzoni 2007; Roelofsen et al. 2019). In contrast, drawing on facts about anaphoricity, focus-sensitivity, and existential presuppositions, we explain this contrast in terms of 'question-to-cleft reduction'.

The rest of this paper is organized as follows. Section 2 presents three observations setting apart *wh*-questions from *whether*-questions, as well as cognitive factives from emotive factives. Section 3 outlines the core assumptions and motivates our question-to-cleft reduction account. Section 4 formalizes this account with dynamic inquisitive semantics (Roelofsen and Dotlačil 2022). Section 5 concludes.

2 Three Observations

Wh- and *whether*-questions differ in anaphoricity, whereas cognitive and emotive factives differ in focus-sensitivity and existential presuppositions.

*We thank Peter Alrenga, Simon Charlow, Michael Glanzberg, and the reviewers of Amsterdam Colloquium 2022 for helpful comments. All errors are ours.

Anaphoricity. Matrix *wh*-questions support discourse anaphora with *wh*-phrases as antecedents (Comorovski 1996; van Rooy 1998; Groenendijk 1998; Aloni and van Rooy 2002; Haida 2007; Dotlačil and Roelofsen 2018; Li 2021; a.o.): in (3a) the pronoun *him* refers to the person that the addressee likes between Andy and Billy. In contrast, as seen in (3b), matrix *whether*-questions do not support anaphora with *wh*-antecedents, no matter whether they are interpreted as alternative questions (*Do you like Andy or [do you like] Billy?*) or as polar questions (*Do you like [or not like] Andy or Billy?*).

(3) a. [Who]^1 do you like between Andy and Billy? You should work with him$_1$.
b. Do you like [Andy or Billy]^1? #You should work with him$_1$.

Similarly, *wh*-embeddings support anaphora, while *whether*-embeddings do not, as seen in (4).

(4) a. I know [who]^1 you like between Andy and Billy. You should work with him$_1$.
b. I know whether you like [Andy or Billy]^1. #You should work with him$_1$.

Unlike other questions, polar questions only accept direct answers like *Yes*/*No*, which shows that the *wh*-antecedents they introduce are propositional. For example, in (5) the anaphor *so* is licensed by the propositional *wh*-antecedent *you like Andy*. The anaphor *him*, in contrast, has a non-*wh*-antecedent *Andy*, licensed in the same way as how it would be in any non-interrogatives.

(5) [Do you like [Andy]^2?]^1 If so$_1$, you should work with him$_2$.

Focus-sensitivity. Embedding emotive factives are focus-sensitive (Romero 2015): in (6a), focus-marking in the embedded declarative must be consistent with the contrast presented in the given context. In contrast, cognitive factives are not focus-sensitive per se, as seen in (6b).

(6) Context: Jane expected Andy to teach semantics. Then she found out that Billy taught semantics.
a. Jane is surprised that BIlly$_F$ taught semantics/#Billy taught seMANtics$_F$.
b. Jane knows that BIlly$_F$ taught semantics/Billy taught seMANtics$_F$.

In parallel, as shown in (7), *wh*-embeddings in Mandarin with an emotive factive require the *wh*-expression (e.g., *shei* 'who') to be associated with an overt focus marker *shi*, in contrast to those with a cognitive factive.

(7) a. Zhangsan hen jingxi #(**shi**) shei gei ta mai-le hua.
Zhangsan very surprised SHI who give him buy-ASP flower.
Zhangsan is very surprised who bought him flowers.
b. Zhangsan zhidao (**shi**) shei gei ta mai-le hua.
Zhangsan know SHI who give him buy-ASP flower.
Zhangsan knows who bought him flowers.

Existential presupposition. *Wh*-embeddings with an emotive factive are subject to an existential (∃-)presupposition (Roelofsen et al. 2019). In (8a) the continuation *if anyone* is infelicitous because it contradicts the presupposition of the preceding question embedding clause that someone won the game. In contrast, in (8b) the continuation is felicitous, which shows that *wh*-embeddings with a cognitive factive do not trigger an ∃-presupposition.

(8) a. Jane is happy who won the game, #if anyone.
b. Jane knows who won the game, if anyone.

3 A Question-to-Cleft Reduction Account

Section 2 presented three observations contrasting *wh-* and *whether-* questions as well as contrasting cognitive and emotive factives. Based on these observations, we propose a question-to-cleft reduction account, which explains the selectional patterns of the two types of factives. This account has two major ingredients: (i) *wh*Qs introduce a unique ans-dref, while AltQs and PolQs introduce multiple ans-drefs and no ans-dref, respectively; (ii) questions embedded under emotive factives must be reduced into clefts.

We call discourse referents introduced by questions that may serve as short answers to questions 'answer discourse referents (ans-drefs)'. The availability and the number of discourse referents vary by question types. First, *wh*Qs introduce one and only one ans-dref (as in (9a)). This assumption resembles what has been commonly assumed for existential statements in dynamic theories as in (9b), that is, indefinites introduce a unique discourse referent (Karttunen 1977; Groenendijk and Stokhof 1991; Dekker 1993; Veltman 1996; a. o.).

(9) a. Which^1 man came to the party? He$_1$ must be bored.
b. A^1 man came to the party. He$_1$ must be bored.

Second, AltQs introduce more than one ans-dref, similar to disjunctions of referential expressions. Simons (1996) observes that in (10a) a disjunction of two non-specific indefinites allows a *whichever* interpretation of the pronoun *she*, while in (10b) a disjunction of two proper names does not. This contrast, however, does not extend to AltQs: neither (11a) nor (11b) supports anaphora.

(10) a. Either a soprano or an alto will sing. She$_{(=\text{whoever sings})}$ will stand on that platform.
b. Either Jane or Maud will sing. #She$_{(=\text{whoever sings})}$ will stand on that platform.

(11) a. Will [a soprano or an alto]^1 sing? #She$_1$ will stand on that platform.
b. Will [Jane or Maud]^1 sing? #She$_1$ will stand on that platform.

Third, PolQs do not introduce any ans-drefs, since they admit only short answers like *Yes* and *No*. In (12) (repeated from (5)), the pronoun *him* takes *Andy* as its antecedent, which however is not a short answer to the polar question.

(12) Do you like Andy^1? You should work with him$_1$.

As the core assumption of our analysis, we argue that questions must be reduced to clefts when embedded under emotive factives. These clefts state that it is the value of the unique ans-dref that satisfies the topical property (Chierchia and Caponigro 2013) of the embedded question. For example, *Jane is surprised who^d won* means that Jane is surprised that it is d (the unique ans-dref introduced by *who*) that won.

This reduction assumption is motivated by the aforementioned facts on focus-sensitivity and ∃-presupposition in embeddings with an emotive factive. Clefts are typical focus constructions, as exemplified in (13). Moreover, clefts trigger an ∃-presupposition: in (14b), the use of *nobody* is odd because it contradicts the cleft's ∃-presupposition (Aravind et al. 2018).

(13) a. It was [Andy's sister]$_F$ who won the lottery.
→ No other people won the lottery.
b. It was [Andy's]$_F$ sister who won the lottery.
→ Nobody else's sister won the lottery.

(14) a. It is Andy who won the lottery. ⇝ *Someone won the lottery.*
b. *It's nobody that won the lottery.

Note that our reduction assumption diverges from Fox's (2020) Question–Cleft Matching Principle, which mandates questions in all types of embeddings to be reduced to clefts. In contrast to Fox, we restrict the mandatory question-to-cleft reduction to only embeddings with emotive factives, based on the fact that embeddings with cognitive factives do not exhibit focus-sensitivity (see (6b) and (7b)) or existential presuppositions (see (8b) and (15)). Moreover, the optionality of reduction in embeddings with cognitive factives is supported by (15): reducing the embedded question in (15a) to a cleft would yield an ungrammatical sentence like (15b), where the ungrammaticality comes from the incompatibility between a *no*-NP and a cleft.

(15) Context: No stocks went up, and Fred knows that.
   a. Fred knows which of the three stocks went up.
   b. *Fred knows that it is no stocks that went up.

With the two assumptions argued above, we are now ready to explain why emotive factives admit *wh*-complements but not *whether*-complements. In short, forming a cleft meaning from an interrogative complement requires to extract a unique ans-dref. This uniqueness requirement is satisfied by *wh*-complements but not by alternative or polar *whether*-complements.

4 Formalizations

Our question-to-cleft reduction account requires a formal theory that captures anaphora across question types. Many attempts have been made to bridge question semantics and anaphoricity, especially in dynamic theories (Jäger 1996; Ginzburg 1996; Hulstijn 1997; Erteschik-Shir 1997; Groenendijk 1999; Aloni and van Rooy 2002; Li 2021; Roelofsen and Dotlačil 2022; a.o.). We follow the framework of dynamic inquisitive semantics developed by Roelofsen and Dotlačil (2022), which captures anaphora in declaratives and all relevant question types.

4.1 Formal Background

In the framework of dynamic inquisitive semantics by Roelofsen and Dotlačil (2022), sentence meanings are composed of atomic '**context update functions (cufs)**'. Cufs can be conjoined via function composition: as in (16a), conjoining two cufs (viz., $\mathcal{U}_1$ and $\mathcal{U}_2$) means applying them to the input context c in sequence. Cufs can also be disjoined, as in (16b), by taking the union of the output contexts yielded by applying each cuf to the input context.

(16) a. $\mathcal{U}_1;\mathcal{U}_2 \coloneqq \lambda c.\mathcal{U}_2(\mathcal{U}_1(c))$
   b. $\mathcal{U}_1 \sqcup \mathcal{U}_2 \coloneqq \lambda c.\mathcal{U}_1(c) \cup \mathcal{U}_2(c)$

Declaratives take the form of a cuf $!\mathcal{U}$, where the !-operator removes any potential inquisitiveness from $\mathcal{U}$. In existential statements, the !-operator applies trivially, since the cuf it scopes over is non-inquisitive. For example, (17) introduces a dref u (represented as a cuf $[u]$) and specifies that u is a person and that u danced. In this formalization, for any predicate R and a dref u, their predication relation is represented as $R\{u\}$, which corresponds to $R(u)$ in the standard predicate logic.

(17) $[\![\textit{Someone danced}]\!] = !([u];\mathsf{person}\{u\};\mathsf{dc}\{u\})$

Roelofsen and Dotlačil assume that disjunctions are inquisitive, unlike existential statements. In (18), the use of the !-operator is non-trivial — it rids the disjunction of its inquisitiveness and renders a non-inquisitive declarative meaning.

(18) $[\![\textit{Either Andy or Billy danced}]\!] = !(([u_1]; \mathsf{A}\{u_1\}; \mathsf{dc}\{u_1\}) \sqcup ([u_2]; \mathsf{B}\{u_2\}; \mathsf{dc}\{u_2\}))$

Interrogatives take the form $\langle ? \rangle \mathcal{U}$. The $\langle ? \rangle$-operator, as defined in (19), ensures the inquisitiveness of questions: it returns a disjunction if the cuf it scopes over is non-inquisitive.

(19) $\langle ? \rangle \mathcal{U} := \begin{cases} \mathcal{U} & \mathcal{U} \text{ is inquisitive;} \\ \mathcal{U} \sqcup \neg \mathcal{U} & \text{otherwise.} \end{cases}$

The *wh*Q in (20) introduces a unique ans-dref u (represented as $\ulcorner u \urcorner$, in contrast to $[u]$),[1] specifies that u is a person and that u danced, and requests u's identity (represented as $?u$). Here the $\langle ? \rangle$-operator applies trivially, since $?u$ is inquisitive.

(20) $[\![\textit{Who danced?}]\!] = \langle ? \rangle(\ulcorner u \urcorner; \mathsf{person}\{u\}; \mathsf{dc}\{u\}; ?u)$

The AltQ in (21) introduces two ans-drefs (viz., u_1 and u_2) and specifies their properties in each disjunct. Here the $\langle ? \rangle$-operator also applies trivially, since disjunctions are inquisitive.

(21) $[\![\textit{Did Andy or Billy dance?}]\!] = \langle ? \rangle((\ulcorner u_1 \urcorner; \mathsf{A}\{u_1\}; \mathsf{dc}\{u_1\}) \sqcup (\ulcorner u_2 \urcorner; \mathsf{B}\{u_2\}; \mathsf{dc}\{u_2\}))$

In PolQs, however, the $\langle ? \rangle$-operator applies non-trivially: it returns a disjunction as the question meaning. Crucially, no ans-dref is introduced; (22) has only a regular/non-ans dref specified to be Andy.

(22) $[\![\textit{Did Andy dance?}]\!] = \langle ? \rangle([u]; \mathsf{A}\{u\}; \mathsf{dc}\{u\}) = ([u]; \mathsf{A}\{u\}; \mathsf{dc}\{u\}) \sqcup \neg([u]; \mathsf{A}\{u\}; \mathsf{dc}\{u\})$

Parthood relations. To extract atomic cufs from a sentence, we define two *part-of* relations, formalized as in (23). Relation (23b) is important for tackling AltQs and PolQs.

(23) a. $\mathcal{U}' \sqsubseteq_c \mathcal{U}$ is true iff $\exists \mathcal{U}_1, \mathcal{U}_2.\ \mathcal{U}_1(\mathcal{U}'(\mathcal{U}_2(c))) = \mathcal{U}(c)$
b. $\mathcal{U}' \sqsubseteq_d \mathcal{U}$ is true iff $\exists \mathcal{U}''.\ \mathcal{U}'(c) \sqcup \mathcal{U}''(c) = \mathcal{U}(c)$
(where $\mathcal{U}_1$,$\mathcal{U}_2$,$\mathcal{U}''$ can be semantically vacuous, i.e., identity functions.)

4.2 Formal Account

First, we define two question-reduction operators, namely, a **QC** (question-to-cleft) operator and a **QP** (question-to-proposition) operator. The **QC** operator reduces a dynamic question to a cleft answer to this question: (a) for any dynamic question $\mathcal{U}$, $\mathbf{QC}(\mathcal{U})$ is defined only if $\mathcal{U}$ introduces a unique ans-dref v; (b) when defined, $\mathbf{QC}(\mathcal{U})$ returns a cleft meaning of $\mathcal{U}$[2], which consists of (i) a presupposition that there exists a unique entity that satisfies all the predicates in $\mathcal{U}$ and (ii) an assertion that the value of this entity is simply the ans-dref introduced by $\mathcal{U}$.

(24) $\mathbf{QC} := \lambda \mathcal{U} : \exists! v. \exists \mathcal{U}' \sqsubseteq_d \mathcal{U}. \ulcorner v \urcorner \sqsubseteq_c \mathcal{U}'.$ (a)
$\lambda w : \exists d. \forall d'. (\forall R. (\exists v'. (R\{v'\} \sqsubseteq_c \mathcal{U}) \rightarrow R(w)(d')) \rightarrow d' \leq d).$ (b-i)
$\forall R. (\exists v. (R\{v\} \sqsubseteq_c \mathcal{U}) \rightarrow R(w)(\mathbf{AD}(\mathcal{U})))$ (b-ii)

The value of the ans-dref is extracted by the **AD**-operator, defined as in (25): $\mathbf{AD}(\mathcal{U})(w)$ returns the unique entity that satisfies all the predicates in $\mathcal{U}$ in the world w.

(25) $\mathbf{AD} := \lambda \mathcal{U}. \lambda w. \iota d. (\forall R. (\exists v. (R\{v\} \sqsubseteq_c \mathcal{U}) \rightarrow R(w)(d)))$

[1] While Roelofsen and Dotlačil (2022) represent discourse referents uniformly as $[u]$, we represent ans-drefs as $\ulcorner u \urcorner$ to distinguish them from other drefs. This paper only makes a notational difference; we need a more refined formal theory to account for the distinctions.

[2] We follow Szabolcsi's (1981) analysis of cleft meaning. Our proposal can be adjusted to adopt other semantic analyses of clefts. See also Bolinger 1972; Atlas and Levinson 1981; Velleman et al. 2012; a.o.

The **QP** reduces a dynamic question to a regular Hamblin answer. Unlike **QC**, **QP** doesn't trigger a unique ans-dref presupposition; $\mathbf{QP}(\mathcal{U})$ simply asserts that the strongest true short answer to $\mathcal{U}$ satisfies all the predicates in $\mathcal{U}$ (if $\mathcal{U}$ is a *wh*Q) or a disjunct of $\mathcal{U}$ (if $\mathcal{U}$ is an AltQ or a PolQ, which denotes a disjunction of question alternatives, as seen in (21) and (22)).

(26) $\mathbf{QP} \coloneqq \lambda\mathcal{U}.\lambda w.\exists\mathcal{U}' \sqsubseteq_d \mathcal{U}.\forall R.(\exists v.(R\{v\} \sqsubseteq_c \mathcal{U}') \to R(w)(\mathbf{SA}(\mathcal{U})))$

The strongest true short answer is extracted by the **SA**-operator, defined as in (27) (c.f. Dayal 1996; Li 2021): $\mathbf{SA}(\mathcal{U})(w)$ returns the unique entity d such that the inferences derived by applying the predicates in (one disjunct of) $\mathcal{U}$ to d are true in w and are the strongest.

(27) $\mathbf{SA} \coloneqq \lambda\mathcal{U}.\lambda w.\iota d.(\exists\mathcal{U}' \sqsubseteq_d \mathcal{U}.(\forall R.(\exists v.(R\{v\} \sqsubseteq_c \mathcal{U}') \to R(w)(d) \wedge \forall d'.(R(w)(d') \to R(d) \subseteq R(d')))))$

Next, we incorporate **QP** and **QC** into the lexical entries of *know* and *be surprised*, as in (28) and (29), respectively (c.f. Uegaki 2015, 2022; Theiler et al. 2018; Fox 2020; a.o.). Both lexical entries make reference to the propositional answer yielded by applying **QP**. *Know* presupposes the truth of this answer and asserts that the attitude holder believes this answer; *be surprised* presupposes the truth of this answer as well as that the attitude holder believes this answer. Crucially, the emotive part of *be surprised* involves the **QC** operator: it asserts that the attitude holder expects the negation of the cleft answer to the embedded question. For example, *Jane is surprised who*^{d} *danced* asserts that Jane previously expected that it isn't d that danced.

(28) $[\![know]\!] = \lambda\mathcal{U}.\lambda x.\lambda w : \mathbf{QP}(\mathcal{U})(w).\textsc{dox}_x^w \subseteq \mathbf{QP}(\mathcal{U})$

(29) $[\![\textit{be surprised}]\!] = \lambda\mathcal{U}.\lambda x.\lambda w : \mathbf{QP}(\mathcal{U})(w) \wedge \textsc{dox}_x^w \subseteq \mathbf{QP}(\mathcal{U}).\textsc{exp}_x^w \subseteq \neg\mathbf{QC}(\mathcal{U})$

As a desirable prediction of the presented account, when a *wh*Q is embedded under *be surprised*, it introduces a unique ans-dref and satisfies the uniqueness presupposition of **QC**; in contrast, an embedded *whether* question introduces either multiple ans-drefs or no ans-dref, causing a presupposition failure to the use of **QC**.

5 Conclusion

This paper investigates why emotive factives do not embed *whether*-questions. First, drawing on evidence from anaphoricity, we argued that different question types vary in the number of ans-drefs they can introduce. Next, motivated by facts on focus-sensitivity and ∃-presuppositions, we argued that questions must be reduced to clefts when embedded under emotive factives. Finally, we formalized our account with the framework of dynamic inquisitive semantics.

References

K. Abels. Why surprise-predicates do not embed polar interrogatives. *Linguistische Arbeitsberichte*, 81:203–222, 2004.

M. Aloni and R. van Rooy. The dynamics of question and focus. In B. Jackson, editor, *Proceedings of Semantics and Linguistic Theory*, volume 12, pages 20–39. CLC Publications, 2002.

A. Aravind, M. Hackl, and K. Wexler. Syntactic and pragmatic factors in children's comprehension of cleft constructions. *Language Acquisition*, 25(3):284–314, 2018.

J. D. Atlas and S. C. Levinson. It-clefts, informativeness and logical form: radical pragmatics (revised standard version). In *Radical Pragmatics*, pages 1–62. Academic Press, 1981.

D. Bolinger. A look at equations and cleft sentences. In *Studies for Einar Haugen: presented by friends and colleagues*, pages 96–114. Mouton & Co. N.V., 1972.

G. Chierchia and I. Caponigro. Questions on questions and free relatives. Presented at the 18th Sinn und Bedeutung, 2013. URL https://scholar.harvard.edu/files/chierchia/files/frs_and_qs_iii_2013-9-11.pdf.

I. Comorovski. *Interrogative phrases and the syntax-semantics interface*. Kluwer Academic, 1996.

V. Dayal. *Locality in WH quantification: questions and relative clauses in Hindi*. Kluwer Academic, 1996.

P.J.E. Dekker. *Transsentential meditations: ups and downs in dynamic semantics*. Doctoral dissertation, University of Amsterdam, 1993.

J. Dotlačil and F. Roelofsen. Dynamic inquisitive semantics: anaphora and questions. In *Proceedings of Sinn und Bedeutung*, volume 23, pages 365–382, 2018.

F.-J. d'Avis. On the interpretation of wh-clauses in exclamative environments. *Theoretical Linguistics*, 28(1):5–31, 2002.

N. Erteschik-Shir. *The dynamics of focus structure*. Cambridge University Press, 1997.

D. Fox. Pointwise exhaustification and the semantics of question embedding. Unpublished Manuscript, 2020.

J. Ginzburg. Interrogatives: questions, facts and dialogue. In S. Lappin, editor, *The Handbook of Contemporary Semantic Theory*, volume 5, pages 359–423. Blackwell Reference, 1996.

J.A.G. Groenendijk. Questions in update semantics. In *Formal Semantics and Pragmatics of Dialogue. Proceedings of Twendial '98*, pages 125–137. University of Twente, 1998.

J.A.G. Groenendijk. The logic of interrogation: classical version. In T. Matthews and D. Strolovitch, editors, *Proceedings of Semantics and Linguistic Theory*, volume 9, pages 109–126, 1999.

J.A.G. Groenendijk and M. Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14 (1):39–100, 1991.

E. Guerzoni. Weak exhaustivity and 'whether': A pragmatic approach. In T. Friedman and M. Gibson, editors, *Proceedings of Semantics and Linguistic Theory*, volume 17, pages 112–129. Cornell University, 2007.

A. Haida. *The indefiniteness and focusing of wh-words*. Doctoral dissertation, Humboldt-Universität zu Berlin, 2007.

J. Hulstijn. Structured information states: Raising and resolving issues. In A. Benz and G. Jäger, editors, *Proceedings of MunDial '97*, pages 99–117, 1997.

G. Jäger. Only updates. In P. Dekker and M. Stokhof, editors, *Proceedings of the Tenth Amsterdam Colloquium*. ILLC, University of Amsterdam, 1996.

L. Karttunen. Syntax and semantics of questions. *Linguistics and Philosophy*, 1(1):3–44, 1977.

H. Li. Mandarin wh-conditionals: a dynamic question approach. *Natural Language Semantics*, 29(3):401–451, 2021.

F. Roelofsen and J. Dotlačil. Wh-questions in dynamic inquisitive semantics. Unpublished Manuscript, 2022.

F. Roelofsen, M. Herbstritt, and M. Aloni. The *whether puzzle. In *Questions in Discourse*, pages 172–197. Brill, 2019.

M. Romero. Surprise-predicates, strong exhaustivity and alternative questions. In S. D'Antonio, M. Moroney, and C. Rose Little, editors, *Proceedings of Semantics and Linguistic Theory*, volume 25, pages 225–245. LSA and CLC Publications, 2015.

K. J. Sæbø. A whether forecast. In B. D. ten Cate and H. W. Zeevat, editors, *International Tbilisi Symposium on Logic, Language, and Computation*, pages 189–199. Springer, 2005.

M. Simons. Disjunction and anaphora. In T. Galloway and J. Spence, editors, *Proceedings of Semantics and Linguistic Theory*, volume 6, pages 245–260, 1996.

A. Szabolcsi. The semantics of topic–focus articulation. In J.A.G. Groenendijk, T.M.V. Janssen, and M.B.J. Stokhof, editors, *Formal Methods in the Study of Language*, pages 513–540. Matematisch Centrum, 1981.

N. Theiler, F. Roelofsen, and M. Aloni. A uniform semantics for declarative and interrogative complements. *Journal of Semantics*, 35(3):409–466, 2018.

W. Uegaki. *Interpreting questions under attitudes*. Doctoral dissertation, Massachusetts Institute of Technology, 2015.

W. Uegaki. *Question-orientedness and the semantics of clausal complementation*. Springer Cham, 2022.

R. van Rooy. Modal subordination in questions. In *Formal Semantics and Pragmatics of Dialogue. Proceedings of Twendial '98*, pages 237–247. University of Twente, 1998.

D. Velleman, D. Beaver, E. Destruel, D. Bumford, E. Onea, and E. Coppock. It-clefts are IT (inquiry terminating) constructions. In A. Chereches, editor, *Proceedings of Semantics and Linguistic Theory*, volume 22, pages 441–460, 2012.

F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3):221–261, 1996.

Recursive numeral systems optimize the trade-off between lexicon size and average morphosyntactic complexity*

Milica Denić[1] and Jakub Szymanik[2]

[1] Tel Aviv University, Tel Aviv, Israel
milica.z.denic@gmail.com
[2] University of Trento, Trento, Italy
jakub.szymanik@gmail.com

Abstract

Human languages vary in terms of which meanings they lexicalize, but there are important constraints on this variation. It has been argued that languages are under the pressure to be simple (e.g., to have a small lexicon size) and to allow for an informative (i.e., precise) communication with their lexical items, and that which meanings get lexicalized may be explained by languages finding a good way to trade off between these two pressures ([12] and much subsequent work). However, in certain semantic domains, it is possible to reach very high levels of informativeness even if very few meanings from that domain are lexicalized. This is due to productive morphosyntax, which may allow for construction of meanings which are not lexicalized. Consider the semantic domain of natural numbers: many languages lexicalize few natural number meanings as monomorphemic expressions, but can precisely convey any natural number meaning using morphosyntactically complex numerals. In such semantic domains, lexicon size is not in direct competition with informativeness. What explains which meanings are lexicalized in such semantic domains? We will propose that in such cases, languages are (near-)optimal solutions to a different kind of trade-off problem: the trade-off between the pressure to lexicalize as few meanings as possible (i.e, to minimize lexicon size) and the pressure to produce as morphosyntactically simple utterances as possible (i.e, to minimize average morphosyntactic complexity of utterances).

1 Introduction

Human languages vary in terms of which meanings they lexicalize (as simple morphemes).[1] There are nonetheless important constraints on this variation: some meanings are frequently lexicalized, while others rarely. For instance, color terms label only convex regions of conceptual color space cross-linguistically [6, 11]. What explains such cross-linguistic generalizations?

A prominent hypothesis is that certain meanings are rarely or not lexicalized because lexicalizing them wouldn't improve communicative efficiency of a language, cf. [12] and much subsequent work. According to this hypothesis, languages are under two pressures: they should be simple (e.g., have simple lexicons), and they should be informative, i.e., allow for a precise communication. These two pressures are in competition. For instance, if a language

*For helpful discussions, we wish to thank Emmanuel Chemla and Roni Katzir, as well as the audiences of the CoSaQ closing workshop in Amsterdam, LINGUAE seminar at École Normale Supérieure in Paris, and Tel Aviv University Linguistics Colloquium. This research was supported by a postdoctoral fellowship awarded to MD by the Azrieli Foundation, and by the funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007- 2013) / ERC Grant Agreement n. STG 716230 CoSaQ.

[1] We consider that a meaning is lexicalized in a language if there is a simple morpheme with this meaning. Whether this morpheme can stand alone as a single word is not part of our criterion for lexicalization. For instance, English lexicalizes the number meaning 6 in a morpheme *six*, and plural meaning in a morpheme *-s*.

only has one word for colors, it will have a simple lexicon, but it won't allow for a very precise communication about colors. Adding more words for various colors would allow for a more precise communication, but at the cost of having a more complex lexicon. Natural languages have been argued to lexicalize the meanings which allow them to achieve a good compromise between these two pressures, i.e., *to optimize the simplicity/informativeness trade-off* [12, 17, 22, 21, 13, 23, 18, 19, 14, 4, 5, 24, 20].

However, in certain semantic domains, it is possible to reach very high levels of informativeness even if very few meanings from that semantic domain are lexicalized. This is due to productive morphosyntax, which may allow for construction of meanings which are not lexicalized. Consider a semantic domain of natural numbers in languages with a so-called *recursive numeral system*: these are languages in which any natural number meaning can be expressed precisely (e.g., English). Many such languages lexicalize few natural number meanings, but can precisely convey any natural number meaning using morphosyntactically complex numerals (e.g., *sixty-one* in English). In such semantic domains, lexicon size is not in direct competition with informativeness. What explains which meanings are lexicalized in these domains?

We will propose that, in semantic domains in which productive morphosyntax enables precise communication even with very few lexicalized meanings, languages are under the pressure to lexicalize as few meanings as possible (i.e, to minimize lexicon size) and the pressure to produce as morphosyntactically simple utterances as possible (i.e, to minimize average morphosyntactic complexity of utterances). Lexicon size and average morphosyntactic complexity of utterances are in competition in such domains: reducing average morphosyntactic complexity of utterances will often require lexicalizing more meanings, and reducing the size of the lexicon will often result in needing utterances of greater morphosyntactic complexity to communicate. *We thus propose that, in such domains, languages lexicalize those meanings which allow them to optimize the trade-off between lexicon size and average morphosyntactic complexity of utterances.*

We will evaluate this proposal within a case study on lexicalized number meanings in languages with a recursive numeral system, and present evidence that recursive numeral systems indeed optimize the trade-off between lexicon size and average morphosyntactic complexity of numerals. This conclusion is in tension with [21], who have instead argued that recursive numeral systems optimize the simplicity/informativeness trade-off — we briefly review their results and explain why we do not think they support the conclusion that recursive numeral systems optimize the simplicity/informativeness trade-off in Appendix B.

Once our results are in place, we will abstract away from numeral systems and discuss which notion of communicative efficiency may explain what meanings get lexicalized across semantic categories more generally, in light of this result and previous work within the simplicity/informativeness trade-off optimization approach [12].

Finally, we will turn to several other lines of work [25, 16, 8, 14, 1] which share with the present work the idea that speakers attempt to minimize complexity of their utterances, and explain in what way they are different from the proposal we are pursuing.

2 Case study on recursive numeral systems

We will present an experiment investigating how close natural languages' recursive numeral systems are to trading off optimally between lexicon size and average morphosyntactic complexity of numerals. We will evaluate this by comparing natural languages to artificially generated languages. The artificial languages represent the space of possibilities in terms of lexicon size/average morphosyntactic complexity of numerals trade-off in recursive numeral systems, and will reveal what the (approximately) optimal solutions to the trade-off problem are.

2.1 Natural languages

We collected cross-linguistic data on numerals denoting numbers 1-99 and their morphosyntactic components in recursive numeral systems from the sample of languages in the *Numeral bases* chapter in *The World Atlas of Language Structures (WALS)* [2]. Out of 172 recursive numeral systems in [2], 44 were excluded due to challenges with data collection or data interpretation. 128 languages were thus included in the analysis.[2]

These numeral systems differ in terms of which number meanings they lexicalize. Some recurring options are listed in (1). The generalization seems to be that they lexicalize the following numbers from the range 1-99: (i) the first n numbers, with n varying across languages, often the first five or the first 10, and (ii) a couple of additional numbers such as 10 and/or 20 as in (1-b) or (1-c).

(1) Lexicalized number concepts:
   a. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 (75 languages)
   b. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20 (19 languages)
   c. 1, 2, 3, 4, 5, 10 (9 languages)

For each of the 128 languages, for each numeral denoting a number in the range 1-99, its morphosyntactic components and their denotations were identified (cf. Table 1 for a few examples of numerals in Fulfulde). The morphosyntactic components of numerals include number-denoting morphemes, and morphemes for arithmetic operators (addition, multiplication and subtraction), cf. [9].[3] Both types of morphemes can in principle be phonetically overt or covert: in practice, number-denoting morhemes are very rarely covert, while morphemes denoting arithmetic operators often are.

Finding morphosyntactic components of numerals required studying the morphosyntactic patterns for numerals in each of the 128 languages. In most languages, the relevant morphosyntacic patterns were described in the descriptive grammars we consulted, or were otherwise easy to detect. There were however more difficult cases, even in languages we (the authors) are closely familiar with (for instance, is there a phonetic variant of the morpheme for number 2 in English *twelve*?). In such cases, we applied the following decision rule: if a numeral is an exception to an established morphosyntactic pattern in a language, but shows (phonetic or orthographic, depending on the available data) elements of morpheme(s) which should have been there if the morphosyntactic pattern was respected, we assume that the numeral follows the morphosyntactic pattern but with phonetic variants not seen elsewhere; otherwise we assume that the numeral is monomorphemic. For instance, English is a base-10 language and because of this we may expect that the numeral for 12 will be built from morphemes for 2 and 10; as the numeral *twelve* has elements of other phonetic realizations of the morpheme for the number 2 (e.g., *two*, *twe(n?)-* from *twenty*), we assume that English *twelve* incorporates the morpheme for number 2 (*twe-*?) and 10 (*-lve*?). Of course, one could argue for the application of an alternative decision rule (unless the morphosyntactic pattern is transparent, assume that the numeral is monomorphemic). It turns out that in practice applying this alternative decision wouldn't qualitatively alter the results; cf. footnote 4.

[2]Natural languages data, including source references for each language, and scripts used for the analysis are available at https://github.com/milicaden/numerals_ac2022.

[3]In the Danish numeral system, a morpheme denoting the fraction *half* is used to construct certain complex numerals. This suggests that there is a limited use of division too involved in the composition of numerals. As, to our knowledge, the extent to which division can be used productively in numeral systems is not well understood [9], we excluded Danish from our corpus.

Table 1: Fulfulde numerals for numbers 6, 30 and 62

Denotation (numeral)	**Morphosyntactic make-up**	**Number of morphemes**
6 (*jowe go'o*)	5 (*jowe*) + (∅) 1 (*go'o*)	3
30 (*chappande tatti*)	10 (*chappande*) · (∅) 3 (*tatti*)	3
62 (*chappande jowe go'o i didi*)	10 (*chappande*) · (∅) (5 (*jowe*) + (∅) 1 (*go'o*)) + (*i*) 2 (*didi*)	7

2.2 Lexicon size and average morphosyntactic complexity

We consider that lexicon size of a language is the number of lexicalized meanings — rather than the number of form-meaning pairs, e.g., we consider that English has one lexicon entry for 10 which can be phonetically realized as *ten*, *-teen*, *-ty*. This is important to keep in mind as it affects how our results should be interpreted: we will argue that languages optimize the trade-off between the number of lexicalized meanings and average morphosyntactic complexity of utterances, rather than the trade-off between the number of form-meaning pairs and average morphosyntactic complexity of utterances. Of course, an important question that remains open is why languages have different phonetic realizations of a single meaning.

Average morphosyntactic complexity of numerals in a language L is computed according to the formula in (2). In (2), $ms_complexity(n, L)$ is the morphosyntactic complexity of the numeral (i.e., the number of morphemes in it) of the language L denoting the number n and $P(n)$ is the probability that the number n needs to be communicated. For instance, average morphosyntactic complexity of English is obtained as $P(1)$ × the number of morphemes in 'one' + $P(2)$ × the number of morphemes in 'two' +...+ $P(99)$ × the number of morphemes in 'ninety-nine'. We assume that the probabilities that different numbers need to be communicated follow a power-law distribution as in (3) (cf. [3, 15, 21]). Qualitatively, this probability distribution captures that the larger the number n, the lower the need to talk about it.

(2)
$$average_ms_complexity(L) = \sum_{n \in [1,99]} P(n) \cdot ms_complexity(n, L)$$

(3)
$$P(n) \propto n^{-2}$$

2.3 Results

In order to evaluate whether natural languages with a recursive numeral system optimize the trade-off between lexicon size and average morphosyntactic complexity of numerals, we need to establish how close they are to the optimal solutions to the lexicon size/average morphosyntactic complexity of numerals trade-off problem. The set of optimal solutions — called *the Pareto frontier* — is a set of (theoretically possible) languages for which there is no other (theoretically possible) language which is better on one of the two dimensions (lexicon size, average morphosyntactic complexity of numerals) without being worse on the other.

We estimated the Pareto frontier using an evolutionary algorithm (cf. [18, 19, 4, 5]), which involves generations of many artificial languages (i.e., artificial recursive numeral systems). We describe this procedure in Appendix A.

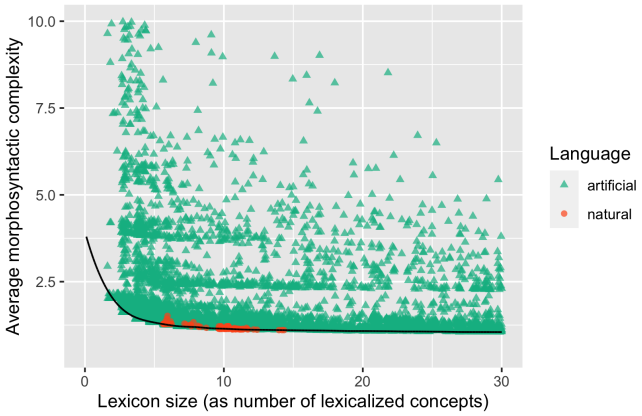


Figure 1: Experiment: Lexicon size and average morphosyntactic complexity of natural languages compared to artificial languages generated via an evolutionary algorithm. Natural languages lie at or very close to the Pareto frontier (black curve).

The natural languages and the artificial languages generated through all generations of the evolutionary algorithm are plotted in Figure 1. The estimated Pareto frontier is plotted as the black curve in Figure 1. Natural languages all lie along or very close to the Pareto frontier in Figure 1. This speaks in favor of natural languages' recursive numeral systems optimizing the trade-off between lexicon size and average morphosyntactic complexity of numerals.[4]

3 Discussion

Three pressures Previous work has provided evidence that in many semantic domains languages lexicalize those meanings which allow them to optimize the simplicity/informativeness trade-off [12, 17, 22, 21, 13, 23, 18, 19, 14, 4, 5, 24, 20].

The semantic domain of natural numbers in languages with a recursive numeral system is an example of a domain in which productive morphosyntax enables precise communication even with very few lexicalized meanings. These are domains in which simplicity of the lexicon is not in direct competition with informativeness. Our results about recursive numeral systems suggest that, in such domains, languages lexicalize those meanings which allow them to optimize the trade-off between lexicon size and average morphosyntactic complexity of utterances. Our work in combination with previous work thus evidences that there are (at least) three pressures shaping lexicons of natural languages: minimizing lexicon size, maximizing informativeness, and minimizing average morphosyntactic complexity of utterances. It is plausible that in many semantic domains the three pressures play a role. More specifically, we expect the three

[4] Recall that, when the number of morphemes in a numeral was difficult to establish, we applied the following decision rule: if a numeral is an exception to an established morphosyntactic pattern in a language, but shows (phonetic or orthographic, depending on the available data) elements of morpheme(s) which should have been there if the morphosyntactic pattern was respected, we assume that the numeral follows the morphosyntactic pattern but with phonetic variants not seen elsewhere; otherwise we assume that the numeral is monomorphemic. The alternative would have been to treat all difficult cases as monomorphemic. Importantly, however, applying the alternative decision rule (treat all difficult cases as monomorphemic) wouldn't alter the results qualitatively. With such an alternative decision rule, natural languages would have slightly larger lexicon sizes and slightly lower average morphosyntactic complexity. In other words, they would be shifted slightly down-right-wards in Figure 1. It is easy to see that they would remain close to the Pareto frontier under such a shift.

pressures to be shaping languages' lexicons in domains in which morphosyntactically complex expressions can improve how precisely meanings from that domain can be communicated, but nonetheless cannot ensure that all the meanings can be communicated precisely.

Previous work on utterance complexity Multiple studies have found that meanings that are conveyed more frequently tend to be associated to shorter forms (with length often operationalized as the number of phonemes) [25, 16, 8, 14]. This is an instance of Zipf's principle of least effort [25], according to which humans are prone to spending the least amount of effort to accomplish a task. The pressure to minimize average morphosyntactic complexity of utterances can be viewed as another instance of the same principle. The novelty of our proposal is in showing that this pressure plays a role in determining which meanings get lexicalized (and not only how long forms are associated with lexicalized meanings). According to our proposal, this pressure on its own doesn't suffice to explain which meanings are lexicalized across languages: we have shown that languages lexicalize meanings which allow them to optimize the trade-off between this pressure and the pressure to minimize lexicon size.

Another recent proposal [1] gives a central role to utterance complexity in explaining typological patterns in the semantic domain of Boolean connectives (*and, or, nor...*). [1] propose that the systems of Boolean connectives are optimal solutions to the trade-off problem between two pressures. The first pressure is the desire to minimize production effort when expressing observations (i.e., *average morphosyntactic complexity*). The second pressure, labeled *conceptual complexity* in [1], relates to how sentences with connectives are used to update the information shared by conversational participants (= *context*). [1] propose that languages are under the pressure to minimize the total conceptual complexity of contextual updates of the connectives they lexicalize. Our proposal shares with [1] the idea that languages are under the pressure to minimize average morphosyntactic complexity. Where the two proposals differ is what the competing pressure is: in our case, the pressure to minimize how many meanings are lexicalized, while in [1] the competing pressure relates to another aspect of language use, namely, how conversational participants update contextual information when they hear an expression.

4 Conclusion

In this paper, we ask what explains which meanings get lexicalized across languages. We pursue the explanation according to which languages lexicalize meanings which allow them to support efficient communication. We however argue for a refinement of what it means for a language to support efficient communication. In particular, the standard approach to communicative efficiency — the simplicity/informativeness trade-off optimization approach — cannot explain which meanings get lexicalized in semantic domains in which lexicon size and informativeness are not in direct competition. We have argued that in such domains languages optimize the lexicon size/average morphosyntactic complexity of utterances trade-off.

Our work in combination with previous work thus evidences that there are (at least) three pressures shaping lexicons of natural languages: minimizing lexicon size, maximizing informativeness, and minimizing average morphosyntactic complexity of utterances. A more general proposal for how supporting efficient communication shapes which meanings get lexicalized across languages may thus be that languages lexicalize those meanings which allow them to solve the trade-off problem between these three pressures in a (nearly) optimal way.

References

[1] Fausto Carcassi and Giorgio Sbardolini. Assertion, denial, and the evolution of boolean operators. *Mind & Language*, 2022.

[2] Bernard Comrie. Numeral bases. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

[3] Stanislas Dehaene and Jacques Mehler. Cross-linguistic regularities in the frequency of number words. *Cognition*, 43(1):1–29, 1992.

[4] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Complexity/informativeness trade-off in the domain of indefinite pronouns. In *Semantics and linguistic theory*, volume 30, pages 166–184, 2021.

[5] Milica Denić, Shane Steinert-Threlkeld, and Jakub Szymanik. Indefinite pronouns optimize the simplicity/informativeness trade-off. *Cognitive Science*, 2022.

[6] Peter Gärdenfors. *The Geometry of Meaning*. The MIT Press, 2014.

[7] Harald Hammarström. *Rarities in numeral systems*. De Gruyter Mouton, 2010.

[8] Martin Haspelmath. Explaining grammatical coding asymmetries: Form–frequency correspondences and predictability. *Journal of Linguistics*, 57(3):605–633, 2021.

[9] James R Hurford. *The linguistic theory of numerals*, volume 16. Cambridge University Press, 1975.

[10] James R Hurford. A performed practice explains a linguistic universal: Counting gives the packing strategy. *Lingua*, 117(5):773–783, 2007.

[11] Gerhard Jäger. Natural Color Categories Are Convex Sets. In Maria Aloni, Harald Bastiaanse, Tikitu de Jager, and Katrin Schulz, editors, *Logic, Language, and Meaning: Amsterdam Colloquium 2009*, pages 11–20. 2010.

[12] Charles Kemp and Terry Regier. Kinship categories across languages reflect general communicative principles. *Science*, 336(6084):1049–1054, 2012.

[13] Charles Kemp, Yang Xu, and Terry Regier. Semantic typology and efficient communication. *Annual Review of Linguistics*, 4:109–128, 2018.

[14] Francis Mollica, Geoff Bacon, Noga Zaslavsky, Yang Xu, Terry Regier, and Charles Kemp. The forms and meanings of grammatical markers support efficient communication. *Proceedings of the National Academy of Sciences*, 118(49), 2021.

[15] Steven T Piantadosi. A rational analysis of the approximate number system. *Psychonomic bulletin & review*, 23(3):877–886, 2016.

[16] Steven T Piantadosi, Harry Tily, and Edward Gibson. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529, 2011.

[17] Terry Regier, Charles Kemp, and Paul Kay. Word meanings across languages support efficient communication. In Brian MacWhinney and William O'Grady, editors, *The Handbook of Language Emergence*, volume 87, pages 237–263. Wiley-Blackwell, Hoboken, NJ, 2015.

[18] Shane Steinert-Threlkeld. Quantifiers in natural language optimize the simplicity/informativeness trade-off. In *Amsterdam Colloquium 2019*, 2019.

[19] Shane Steinert-Threlkeld. Quantifiers in natural language: Efficient communication and degrees of semantic universals. *Entropy*, 23(10):1335, 2021.

[20] Wataru Uegaki. The informativeness/complexity trade-off in the domain of boolean connectives. *Linguistic Inquiry*, pages 1–39, 2022.

[21] Yang Xu, Emmy Liu, and Terry Regier. Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Open Mind*, 4:57–70, 2020.

[22] Yang Xu, Terry Regier, and Barbara C. Malt. Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40(8):2081–2094, 2016.

[23] Noga Zaslavsky, Charles Kemp, Terry Regier, and Naftali Tishby. Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, 115(31):7937–7942, 2018.

[24] Noga Zaslavsky, Mora Maldonado, and Jennifer Culbertson. Let's talk (efficiently) about us: Person systems achieve near-optimal compression. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, 2021.

[25] George Kingsley Zipf. *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA, 1949.

A Evolutionary algorithm

[9, 10] in his seminal work proposes that number-denoting morphemes across languages are of the syntactic category Digits (D) or Multipliers (M). Roughly, morphemes denoting numbers which are 'bases' as 10 or 20 in 'base-10' or 'base-20' languages are of the syntactic category M; other morphemes are of the syntactic category D. He proposes that, across languages, numerals are constructed according to the grammar in (4). We will use this grammar to generate numerals of artificial languages.

(4) NUMBER $\longrightarrow$ D | PHRASE | PHRASE + NUMBER | PHRASE − NUMBER
PHRASE $\longrightarrow$ M | NUMBER · M

For our purposes, an artificial language consists of (i) a lexicon of number-denoting morphemes, which are of category D or M (cf. the grammar in (4)), and (ii) a set of numerals for numbers 1-99 generated according to the morphosyntactic rules in (4) from the lexicon, where each numeral is the shortest expression (or one of them, in case of a tie) for a given number. For example, an artificial language L may have in its lexicon morphemes for 1, 2, 4, 6, 8, 10, such that 1, 2 and 4 are of category D and 6, 8 and 10 are of category M. The numeral for, for instance, number 16 in L would be randomly selected from one of the three morphosyntactically simplest options: $10 + 6$, $2 \cdot 8$, $8 + 8$ ($4 \cdot 4$ is not an option because it cannot be generated by the grammar in (4) (because 4 is of category D in L).

Due to computational constraints, two restrictions are imposed on the search for the shortest expression for a given number in a language. (i) Expressions of depth x (i.e., expressions with at most x number-denoting morphemes and $x - 1$ arithmetic operators) are incrementally constructed from expressions of lower depths (e.g., expressions of depth 2 are constructed by combining expressions of depth 1). However, at all depths, the meaning of expressions is restricted to be a natural number in $[1, 200]$. (ii) If no expression for a number meaning is found with a depth of at most 7, the search is abandoned, and the language is discarded.

The evolutionary algorithm works as follows. First, the generation 0 is created, which consists of 2000 artificial languages. The lexicons of these artificial languages are generated by drawing two random samples of numbers between 1-99; these stand for morphemes of category D and M respectively. As natural languages tend to have very few morphemes of category M (often only 1 or 2), we restrict the size of the random sample for the category M to at most 5 (no such restriction is imposed on the category D). The numerals for numbers 1-99 of these languages are generated as explained above. The dominant languages of a generation (those for which there is no language which is better on one dimension of lexicon size and average morphosyntactic complexity without being worse on the other) each give rise to an equal number of offspring languages, which are obtained via a small number of mutations (between 1 and 3; these mutations included removing a number morpheme from D or M, adding a number morpheme to D or M, and interchanging a number morpheme in D or M, making sure that

M is no larger than 5) from dominant languages. The dominant languages from generation 0 together with their offspring languages constitute generation 1, whose size is limited to 2000 languages. This process is repeated for 100 generations. Finally, the dominant languages are selected from the union of the last generation and the natural languages. Each of these dominant languages is a point in a two-dimensional (lexicon size and average morphosyntactic complexity of utterances) space; we do spline interpolation of these points to form the Pareto frontier.

B Simplicity/informativeness trade-off in numeral systems [21]

[21] analyze the simplicity/informativeness trade-off in 24 *restricted* and 6 *recursive* numeral systems. Restricted numeral systems have numerals for only the first few numbers, and use a quantifier such as *many* for any higher number. For instance, the language Krenak only has numerals for numbers 1-3 [7]. Recursive numeral systems are considered by [21] to be maximally informative when it comes to communicating about number meanings, while restricted numeral systems have lower degrees of informativeness. On the other hand, according to their approach to measuring complexity, the six studied recursive numeral systems are more complex than most of the 24 studied restricted numeral systems. Simplifying somewhat, this is because they assume that the complexity measure of a language should incorporate both the complexity of the lexicon and the complexity of morphosyntactic rules. As recursive numeral systems always have morphosyntactic rules for building numerals but restricted numeral systems often don't have any, recursive numeral systems tend to have a greater measure of complexity than restricted numeral systems. Recursive numeral systems are thus more complex and more informative than restricted numeral systems in [21]. [21] conclude that natural languages' numeral systems optimize the simplicity/informativeness trade-off, without making a distinction between restricted and recursive languages.

Their conclusion may seem to be in tension with our findings. However, a careful examination of the results in [21] reveals that, while restricted numeral systems are indeed close to being Pareto-optimal in trading off simplicity and informativeness, recursive numeral systems are not. Given that recursive numeral systems are maximally informative, if they were optimizing the simplicity/informativeness trade-off, we would expect to find them in the proximity of the minimally complex numeral system which is maximally informative. According to the results of [21] (cf. Figure 4b in [21]), the maximally informative system with minimal complexity has complexity ≈ 50. Strikingly, the six recursive numeral systems examined by [21] have much higher complexity than that, ranging between ≈ 75 and ≈ 150. The results of [21] thus also do not support the hypothesis that recursive numeral systems optimize the simplicity/informativeness trade-off — in other words, there is no tension between the results of [21] and our findings. Importantly, however, conclusions of [21] still hold for restricted numeral systems which are much closer to being Pareto-optimal than recursive numeral systems.

An experimental study on kind and generic readings across languages: bare plural vs. definite plural*

Imke Driemel[1†], Johannes Hein[1‡], Desiré Carioti[2§], Jakob Wünsch[3¶], Vina Tsakali[4‖], Artemis Alexiadou[3,1**], Uli Sauerland[3††] and Maria Teresa Guasti[2‡‡]

[1] Humboldt University, Berlin, Germany
imke.driemel@hu-berlin.de, johannes.hein@hu-berlin.de
[2] University Milano-Bicocca, Milan, Italy
desire.carioti@unimib.it, mariateresa.guasti@unimib.it
[3] Leibniz ZAS, Berlin, Germany
wuensch@leibniz-zas.de, sauerland@leibniz-zas.de, artemis@leibniz-zas.de
[4] University of Crete, Crete, Greece
tsakali@uoc.gr

Abstract

To express kind and generic readings, Romance languages like Italian have been reported to use definite plurals but Germanic languages like English make use of bare plurals (Krifka et al., 1995; Chierchia, 1998), where Greek patterns with Romance (Alexiadou et al., 2007). German is discussed as an exception, as both bare and definite plurals are used to express kind (Krifka et al., 1995) and generic (Longobardi, 1994) readings. We present results from an experimental study comparing English, German, Italian and Greek, focusing on two novel findings: i) German and English both express kind/generic readings with bare plurals; ii) the effect of speaker distance (Acton, 2019) makes the definite plural an additional option in German.

1 Background

Generic and kind readings can be realized by a multitude of structures within and across languages. Italian and Greek, e.g., use definite plurals but English makes use of bare plurals (Krifka et al., 1995; Chierchia, 1998; Alexiadou et al., 2007), compare (1) to (2)/(3).

(1) English (Chierchia, 1998)

a. (*The) dogs are rare. *kind*

b. (*The) dogs love to play. *generic*

*We thank Natale Stucchi for pointing us to the Thurstone Method. Ethical review of the experimental protocols was done by the review boards of the University of Milan-Bicocca (RM-2020-229), the German Linguistic Society (#2019-20A-220622), and the ERC Executive Agency. Consent was obtained from participants. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (LeibnizDream, grant agreement No 856421).

†Contributions: Conceptualization, Methodology, Resources, Writing – Original Draft.
‡Contributions: Conceptualization, Methodology, Resources.
§Contributions: Resources, Methodology, Formal Analysis, Writing – Original Draft.
¶Contributions: Resources, Methodology.
‖Contributions: Resources.
**Contributions: Conceptualization, Resources, Funding acquisition.
††Contributions: Conceptualization, Resources, Funding acquisition.
‡‡Contributions: Conceptualization, Resources, Methodology, Funding acquisition.

(2) Italian (Chierchia, 1998)

a. *(I) cani sono rari. *kind*
the dogs are rare
'Dogs are rare.'

b. *(I) cani amano giocare. *generic*
the dogs love to.play
'Dogs love to play.'

(3) Greek (Lazaridou-Chatzigoga and Alexiadou, 2019; Alexiadou et al., 2007)

a. *(Ta) pulja dodo ehun pleon afanisti. *kind*
the birds dodo have already disappeared
'Dodo birds have already disappeared.'

b. *(I) ghates ine aksiolatrefta plasmata. *generic*
the cats are adorable creatures
'Cats are adorable creatures.'

German is reported to be exceptional, as both definite plurals and bare plurals license generic/kind readings, shown in (4).

(4) German (Krifka et al., 1995; Longobardi, 1994)

a. (Die) Pandabären sind vom Aussterben bedroht. *kind*
the pandas are from extinction facing
'Pandas are facing extinction.'

b. (Die) Bieber bauen Dämme. *generic*
the beavers build dams
'Beavers build dams.'

To account for (1) vs. (2/3), Chierchia (1998) proposes that arguments in English are mapped to kinds, formed by the kind operator $\cap$, a function from worlds to the sum of all instances of the kind in that world. Arguments in Italian (and Greek one can argue), however, are mapped to properties, thus they receive a kind reading via $^\wedge\iota$, see (5). Since the definite determiner realizes ι and not $\cap$ and the Blocking Principle enforces overt over covert type-shifting, Italian/Greek kind nouns are always overtly marked for definiteness, in contrast to English. Moreover, only plural nouns are licensed since a property holding of only one individual per world is not a natural kind.

(5) a. English (1a): $rare(\cap(dogs))$

b. Italian (2a): $rare(^\wedge\iota(dogs))$

Dayal (2004) adopts Chierchia's semantics in (1a) for Germanic, and crucially also for Romance languages. To account for the cross-linguistic split, she then proposes the ranking $\iota > \cap$ where the definite determiner may lexicalize either both ι and $\cap$, as in Romance, or only ι, as in English. The benefit of Dayal's approach lies in accounting for German, where it is argued that both ι and $\cap$ are lexicalized but the Blocking Principle is assumed to be inactive in the kind/generic domain, leading to the observed optionality in (4).

The observations regarding kinds can also be found in the generic domain since generic readings are often argued to be built on kinds, though they crucially also involve a Gn operator (Carlson, 1977; Krifka et al., 1995). As shown in (6a), Chierchia (1998) uses Gn to analyze generic readings as quantification over situations, where situations are contextually restricted. In prose, (6a) expresses that for every individual x and situation s such that x is a dog and s stands in relation C to x, x loves to play in s. For Italian, Chierchia proposes logically equivalent structures without making use of $\cap$, see (6b).

(6) a. English (1b): Gn x, s $[^{\cup\cap}\text{dog}(x) \wedge C(x, s)]$ $[love.play(x, s)]$

b. Italian (2b): Gn x, s $[x \leq \iota\text{dogs} \wedge C(x, s)]$ $[love.play(x, s)]$

The empirical facts about English, however, have recently been challenged. Acton (2019) observed that generic readings can be expressed by definite plurals in English after all, though with the effect that the speaker distances themselves from the kind expressed, shown in (7). A natural question that arises is whether the same effect exists for other Germanic languages like German and if this effect could be one of the reasons for the perceived optionality between bare plurals and definite plurals in (4).

(7) The distance effect (Acton, 2019)

a. Americans love cars.

b. The Americans love cars. $\rightsquigarrow$ *speaker distance*

We investigate the introspective judgements reported in the literature with a quantative study, aiming to tease apart kind and generic readings as well as the additional distance effect shown in (7).

2 Experiment

Little experimental work has been done on the contrasts reported in the previous section. Barton et al. (2015) conducted an acceptability judgement study (yes-no task) for German arguing for the optionality in (4). Ionin et al. (2011) used an acceptability judgement rating task (Likert scale) to verify the Germanic–Romance split, compare (1) and (2), and the well defined kind restriction (Carlson, 1977; Dayal, 2004) for English, Spanish and Brazilian Portuguese. Lazaridou-Chatzigoga and Alexiadou (2019) replicate this study for Greek. We present results from an acceptability judgement study that makes use of pairwise forced choice comparison. Unlike previous studies, we investigate the acceptability of sentences using Thurstone's method of paired comparison (Thurstone, 1927a). We chose this task over a yes-no or rating task, as it is often difficult to explicitly accept or reject an utterance, and to quantify the acceptability of a sentence. Thurstone proposed that the choice between the preferred options of a pair can be a useful solution to order options on an objective scale independent from the individual observer (Thurstone, 1927a). The languages investigated for the kind/generic contrasts are English, German, Italian, and Greek. As both definiteness and number seem to be relevant in expressing kinds/generics, we tested 4 different noun types: singular definite, singular indefinite, bare plural, definite plural.

2.1 Participants

581 adult participants aged between 18 and 60 years took part in the study: 152 native speakers of English (age on average = 29.54, SD = 8.83), 155 native speakers of German (age on average = 24.53, SD = 5.1), 122 native speakers of Italian (age on average = 23.41, SD = 5.17), and 152 native speakers of Greek (age on average = 24.55, SD = 8.93). They were recruited online, through the Prolific platform (German, English, Greek) and the SONA System (Italian).

2.2 Materials and Method

Each participant was presented with 9 contexts and 4 sentences as possible completions for each of the contexts. We focus on the results of 3 contexts in this paper: the generic context (8), the kind context (9), and the generic distance context (10).[1] The target sentences presented in each context only differed in that they started with one of the four different nominals (singular definite, singular indefinite, bare plural, definite plural). Note that while we are presenting all 4 options in (8)–(10) for illustrative purposes, participants saw only two options per trial.

(8) Generic context: There are many pests in the world that make our lives difficult. They eat our supplies, disturb our sleep, or plainly get on our nerves. For example:

a. Mosquitos give us itchy bites in the summer.

b. The mosquitos give us itchy bites in the summer.

c. The mosquito gives us itchy bites in the summer.

d. A mosquito gives us itchy bites in the summer.

(9) Kind context: The constant growth of the human population on earth has taken and still is taking its toll on other life on the planet, plant or animal. For example...

a. Pandas are almost extinct.

b. The pandas are almost extinct.

c. The panda is almost extinct.

d. A panda is almost extinct.

(10) Generic distance context: There is a place in town where people meet for a drink and a chat after work. As there are federal elections coming up soon, a lot of the discussions and debates revolve around politics. Yesterday, one guest seemed very upset and continuously complained that "voting is meaningless because ...

a. politicians do whatever they want after the election anyway."

b. the politicians do whatever they want after the election anyway."

c. the politician does whatever s/he wants after the election anyway."

d. a politician does whatever s/he wants after the election anyway."

[1] The other contexts constitute i) controls; ii) a context filtering for a normative reading; iii) two non-well defined kind/generic contexts set up to investigate the well defined kind restriction. For more information, see the preregistration https://osf.io/wzxqf.

For each context, all possible paired combinations of the 4 sentences (6 pairs) appeared successively, resulting in 54 trials per participant. Participants had to indicate their preferred option for each sentence pair.

2.3 Analysis

We used Thurstone scaling to derive a linear rating of all four nominal forms from two-way comparisons. Thurstone's method of paired comparisons (Thurstone, 1927a,b) is a methodological procedure for psychophysical judgments. It is considered the gold standard for collecting, analyzing, and interpreting subjective introspective data (see Montag 2006; Cattelan 2012; Parraga 2015 for reviews). This method allows deriving the scaling of subjects' preferences, quantifying the distance between two options. As a consequence, we obtain not just the preference order but also, from an interpretative perspective, the degree of acceptability of each option compared to the other. Once we obtained the percentage degree in which an option has been preferred in each pair, a z-score has been derived for each option. Between-options distances were interpreted based on z-scores as effect sizes. Here we only report the between-options results. All the analyses were performed in the R software environment for statistical computing (R Core Team, 2022).

2.4 Results

In Fig. 1, we show the results for the kind and the generic context for every language sample, based on the z-scores. Overall, we can see two clusters, English and German on the one hand and Italian and Greek on the other. In both contexts, bare plural was the best alternative for both English- and German-speaking participants, while plural definite was the best choice for Italian- and Greek-speaking participants. Thus, there is a clear distinction between the two sets of languages in the way they express genericity and kind readings.

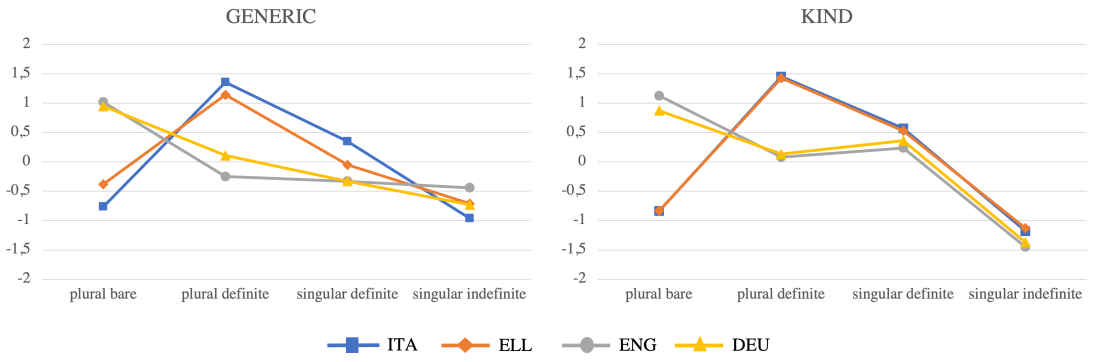


Figure 1: z-scores plotted on y-axis

In the distance context, shown in Fig. 2, German-speaking participants considered bare plural and plural definite equally acceptable. For English-speaking participants in the same context, the plural definite has a lower probability of being chosen than the bare plural, although it still remains the second best option. This means that in the distance context plural definite is a valid option for both languages, and more so in German than in English.

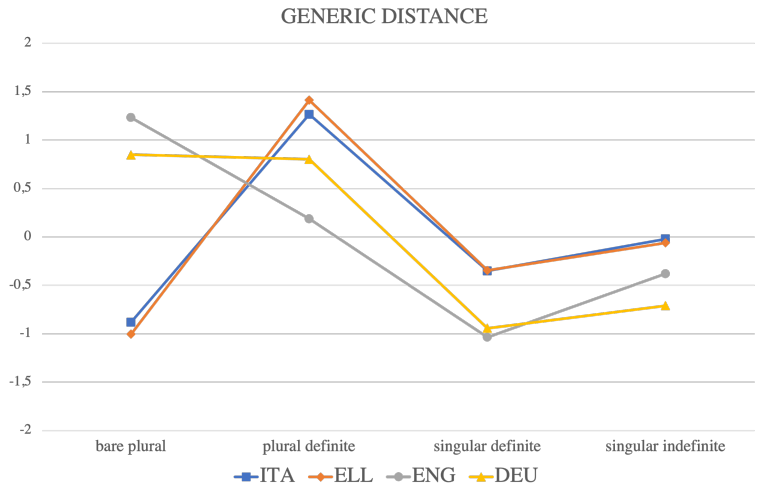


Figure 2: z-scores plotted on y-axis

3 Discussion

The results in Fig. 1 indicate that German is similar to English in expressing kinds/generics with bare plurals. In this sense, German is not exceptional (*pace* Longobardi 1994; Krifka et al. 1995; Dayal 2004) but patterns with English considering kind and generic contexts. This seems to be a welcome result, as the perceived optionality in German for (4) was explained by Dayal (2004) in terms of weakening the Blocking Principle in the kind/generic domain for German. Given that we find a clear distinction between Italian/Greek speakers, who use the definite plural, and English/German speakers, who use the bare plural as a default generic expression, our results support the universality of the Blocking Principle in the sense of Chierchia (1998).

The interesting finding comes from the distance context in Fig. 2 where in particular German speakers make use of definite plurals in addition to bare plurals. As expected by Acton (2019), a context which singles out a reading where speakers presumably distance themselves from the kind expressed in the target sentence boosts the use of the definite plural compared to the other kind/generic contexts shown in in Fig. 1. The default bare plural, however, is also used to an equal amount. This indicates either that only half of the speakers distance themselves from the kind expressed, or, more likely, that the expression of speaker distance is optional. The distance effect was also observed for English albeit to a much lower degree, which comes as a surprise. It could be explained by the fact that, due to ethical requirements, we could only recruit English speakers living in Ireland. Acton's corpus study, however, is based on speeches given in the US House of Representatives, thus representing American English. It is likely that register and/or cultural differences influence the use of distance marking.

References

Acton, E. K. (2019). Pragmatics and the social life of the English definite article. *Language*, 95:37–65.

Alexiadou, A., Haegeman, L., and Stavrou, M. (2007). *Noun Phrase in the Generative Perspective*. Mouton de Gruyter, Berlin.

Barton, D., Kolb, N., and Kupisch, T. (2015). Definite article use with generic reference in German: an empirical study. *Zeitschrift für Sprachwissenschaft*, 34:147–173.

Carlson, G. (1977). *Reference to kinds in English*. PhD thesis, University of Massachusetts, Amherst.

Cattelan, M. (2012). Models for Paired Comparison Data: A Review with Emphasis on Dependent Data. *Statistical Science*, 27:412–433.

Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6:339–405.

Dayal, V. (2004). Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, 27:393–450.

Ionin, T., Montrul, S., and Santos, H. (2011). An experimental investigation of the expression of genericity in English, Spanish and Brazilian Portuguese. *Lingua*, 121:963–985.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meulen, A., Link, G., and Chierchia, G. (1995). Genericity: An introduction. In Pelletier, F. J. and Carlson, G. N., editors, *The Generic Book*, pages 1–124. The University of Chicago Press, Chicago/London.

Lazaridou-Chatzigoga, D. and Alexiadou, A. (2019). Genericity in Greek: an experimental investigation. In Gattnar, A., Hörnig, R., Störzer, M., and Featherston, S., editors, *Proceedings of Linguistic Evidence 2018: Experimental Data Drives Linguistic Theory*, pages 245–260. The University of Tübingen, Tübingen.

Longobardi, G. (1994). Reference and proper names: A Theory of N-movement in Syntax and Logical Form. *Linguistic Inquiry*, 25:609–665.

Montag, E. D. (2006). Empirical formula for creating error bars for the method of paired comparison. *Journal of Electronic Imaging*, 15:010502.

Parraga, C. A. (2015). Perceptual Psychophysics. In Cristóbal, G., Perrinet, L., and Keil, M. S., editors, *Biologically Inspired Computer Vision*, pages 81–108. Wiley.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review*, 34:273.

Thurstone, L. L. (1927b). The method of paired comparisons for social values. *The Journal of Abnormal and Social Psychology*, 21:384.

In search of a unified mechanism for regularisation of linguistic variation

Aislinn Keogh, Simon Kirby and Jennifer Culbertson

Centre for Language Evolution, University of Edinburgh
aislinn.k.keogh@gmail.com

1 Introduction

Natural languages rarely exhibit variation that is not predictable according to some criteria [7]. For example, there is considerable variation in English plural marking, but the choice of variant is lexically-conditioned – so *dog* → *dogs*, but *goose* → *geese* and *fish* → *fish*. Experimental and computational work (e.g. [6, 20, 11, 17]) has demonstrated that individuals have cognitive biases which work against *unpredictable* variation. When no conditioning is present, learners tend to regularise variation (e.g. by increasing use of one rule to the exclusion of others).

Furthermore, even *predictable* variation is sometimes lost as languages evolve, with previously irregular forms (e.g. *cow* → *kine* in Middle English) adopting the regular pattern. Does this kind of language change reveal a similar bias against predictable variation? The most parsimonious explanation is that predictable and unpredictable variation are not two fundamentally different phenomena, but that they exist on one spectrum of complexity, with greater predictability equating to greater simplicity and the simplest possible system being one with no variation at all [3]. If this is the case, the same cognitive biases that act on unpredictable variation should also reduce even predictable variation, albeit potentially less strongly or over a longer time frame. However, there is some experimental evidence that conditioned variation is treated differently from unconditioned variation [9, 11, 18]. Pragmatics-based accounts of regularisation (e.g. [14]) also predict that, while languages will not sustain unpredictable variation, conditioned variation should be evolutionarily stable.

In the present study, we use an artificial language learning experiment to investigate whether the same cognitive biases target both predictable and unpredictable variation. Experimental evidence consistently points to a production-side bias for regularity rather than a learning-based account [6, 14, 11]. More specifically, limitations on memory retrieval during language production have been suggested as one potential mechanism for the loss of *unpredictable* variation [10]. Here, we further test this hypothesis, providing evidence that regularisation of both predictable and unpredictable variation arise via this same mechanism.

We train participants on a language exhibiting either predictable or unpredictable variation in plural marking, and with working memory taxed during learning, production, or not at all. In line with the production-side account of regularisation, we predicted that participants across conditions would show no evidence of having encoded a more regular language than the one they were trained on. In line with the memory retrieval hypothesis, we predicted that we would see the clearest evidence for reduction of variation when taxing working memory during production. Finally, to test our hypothesis regarding variation type, we predicted that the effect of memory limitations during production would be modulated by variation type, with greater regularisation of unpredictable languages.

2 Methods

The study was pre-registered with the Open Science Foundation (https://osf.io/vqyej).

Design: Participants were randomly assigned to one of six conditions in a 2x3 between-subjects design. Their input languages exhibited either probabilistically lexically-conditioned (PREDICTABLE conditions) or random (UNPREDICTABLE conditions) variation. To create a memory pressure, we use an interference task known to specifically tax working memory [13]. This was administered either during training (LEARNING LOAD conditions), production (PRODUCTION LOAD conditions) or not at all (NO LOAD conditions).

Participants: Participants were 220 adult, self-reported native English speakers with no known language disorders, recruited via Prolific and paid £3 for up to 20 minutes' participation. Following our pre-registered exclusion criteria, we excluded 47 participants from analysis for the following reasons: self-reporting the use of written notes in an exit questionnaire contrary to instructions (3), data saving errors (1), failing to provide usable data on more than two critical trials (38), and button mashing (5). This left us with data from 173 participants[1].

Materials: Participants were asked to learn a small artificial language consisting of text labels paired with six images. Each image depicted a pair of animals, and was described by a two-word label: one word for the noun and one word indicating plurality (presented in the English frame 'Here are two...'). Noun labels were iconic (e.g. "buzzo" for a bee) and paired with one of two plural markers, both non-English CVC monosyllables ("mej" and "huv"). The mapping of nouns to plural markers varied according to condition as shown in Table 1.

	Predictable							**Unpredictable**						
	N1	**N2**	**N3**	**N4**	**N5**	**N6**	**Total**	**N1**	**N2**	**N3**	**N4**	**N5**	**N6**	**Total**
P1	7	7	7	7	1	1	30	5	5	5	5	5	5	30
P2	1	1	1	1	7	7	18	3	3	3	3	3	3	18
Total	8	8	8	8	8	8	48	8	8	8	8	8	8	48

Table 1: Distribution of plural markers (P_i) across nouns (N_j) in the two variation conditions. In PREDICTABLE conditions, four nouns were randomly assigned to one plural marker (the 'regulars') and two to the other marker (the 'irregulars'). A small amount of noise was then added to this mapping, such that each noun appeared with its assigned plural marker 87.5% of the time, and with the other 12.5% of the time. This noisy conditioning meant that participants could regularise without having to produce a description they had never observed. In UNPREDICTABLE conditions, all nouns appeared with one marker 62.5% of the time, and with the other 37.5% of the time. Both markers appeared with the same overall frequency in the two variation conditions. The assignment of specific nouns to the N slots and specific plural markers to the P slots was randomised for each participant.

Procedure: The experiment was written in JavaScript using the JsPsych library [5] and ran in participants' web browsers. Participants were randomly assigned to one of the six conditions at the start of the experiment. In all conditions, the experiment consisted of three phases: training, production, and estimation. The order of presentation was randomised within each phase. A schematic of the experiment is given in Figure 1.

On each training trial, participants were shown an image and a corresponding description. Each of the six images was shown eight times for a total of 48 trials. On each production trial, participants saw an image and had to complete the description by clicking two buttons from an array consisting of all nouns and all plural markers in the language. As in training, each of

[1] PREDICTABLE/NO LOAD: 29; PREDICTABLE/LEARNING LOAD: 30; PREDICTABLE/PRODUCTION LOAD: 29; UNPREDICTABLE/NO LOAD: 28; UNPREDICTABLE/LEARNING LOAD: 28; UNPREDICTABLE/PRODUCTION LOAD: 29

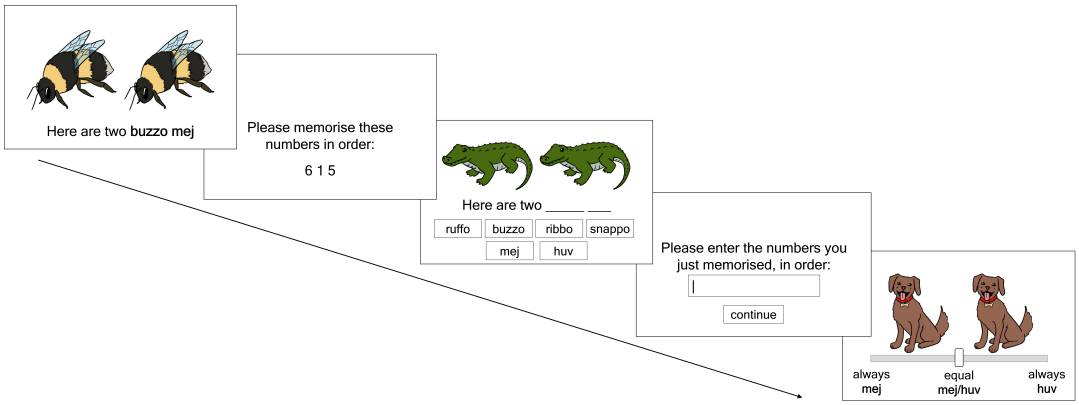


Figure 1: Schematic of the experiment: PRODUCTION LOAD condition. Top to bottom: training trial, digit sequence presentation, production trial, digit sequence recall, estimation trial. Participants in LEARNING LOAD conditions would instead have seen the digit sequence presentation and recall trials sandwiched around each training trial. Participants in NO LOAD conditions would not have seen these digit sequence trials.

the six images was shown eight times for a total of 48 trials. Finally, in the estimation phase, participants were shown all six images and asked to estimate how often they had seen each noun with each plural marker in training. They provided these estimates using a continuous slider with three labels: "always *plural 1*" on the left, "equal *plural 1/plural 2*" in the middle, and "always *plural 2*" on the right.

In LEARNING LOAD and PRODUCTION LOAD conditions, participants additionally completed a digit sequence recall task (modelled after [13]) during the training or production phase, respectively. Before every trial, participants were asked to memorise a pseudo-random sequence of three digits. Immediately after each trial, they were asked to retype the digits. They were given feedback on how many digits they had recalled in the correct position and how long they had taken to respond, to encourage both speed and accuracy.

3 Results

Following previous work [6, 14, 20], we quantify regularisation in information theoretic terms [19]. Here, we measure the entropy of the plural marking system participants estimated and produced, relative to the input language. When entropy decreases, we can infer that one variant has become more frequent overall. The input languages are matched on this measure (0.95 bits). The degree of lexical conditioning – or predictability – of plural marking is given by the mutual information between nouns and plural markers. Again, we calculate this for participants' estimates and productions, relative to the input. When mutual information increases, we can infer that variation has become more strongly lexically-conditioned. Unlike for entropy, the input languages differ in mutual information: UNPREDICTABLE languages score 0 (indicating no lexical conditioning), while PREDICTABLE languages score 0.41 (indicating imperfect lexical conditioning). Our hypothesis is that regularisation of both types of variation arises from a production-side bias driven by memory limitations. If this is the case, we would expect no reliable change in either measure between the input and participants' *estimates*, but a reliable change in at least one of the measures between the input and participants' *productions* in PRODUCTION LOAD conditions, especially for UNPREDICTABLE languages (which are more

complex [3]).[2]

We first investigate regularisation during learning by analysing the change in entropy between participants' training data and the languages described by their estimates, as shown in Figure 2a. We found no reliable decrease in entropy: confidence intervals around the mean for all conditions are either above or crossing zero.

However, as Figure 2b suggests, there was an increase in mutual information between the languages participants in UNPREDICTABLE conditions were trained on and the ones described by their estimates: confidence intervals around the mean are all above zero for these conditions. Pairwise comparisons between conditions further reveal that, while there are no differences between memory load conditions within each variation type, every UNPREDICTABLE condition is reliably different from every PREDICTABLE condition.

To summarise, our results suggest that while overall variation is not reduced during learning, learners *are* biased to infer patterns of conditioning when variation in the input is unpredictable.

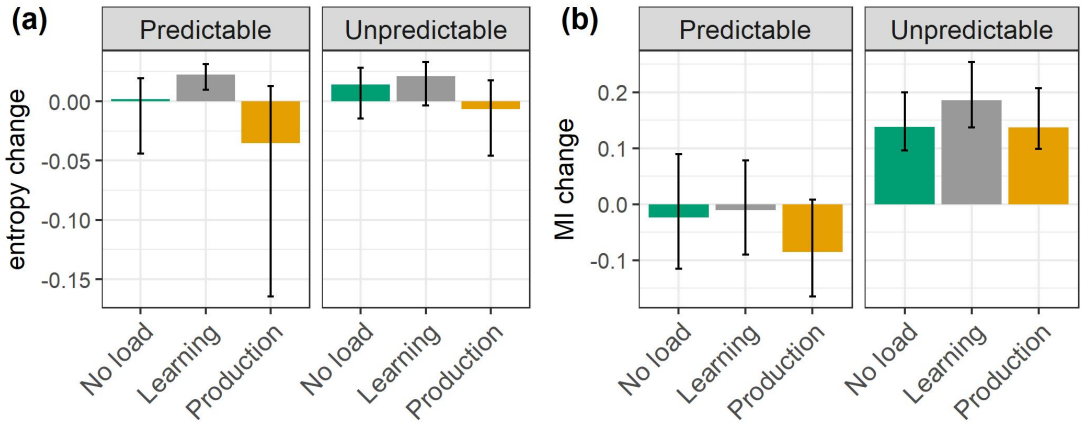


Figure 2: Change in entropy (left) and mutual information (right) from the training data to participants' estimates, by condition. Error bars represent bootstrapped 95% confidence intervals over the mean.

We next investigate regularisation during production by analysing the change in entropy between participants' training data and the languages they produced, as shown in Figure 3a. Here we find evidence for a drop in entropy in PRODUCTION LOAD conditions only (confidence intervals around the mean are below zero for both these conditions, but cross zero for all other conditions). However, we find no evidence that this effect is stronger in the UNPREDICTABLE condition[3].

As suggested by Figure 3b, we observed an increase in mutual information across all conditions except PREDICTABLE/PRODUCTION LOAD: confidence intervals around the mean are

[2]Inspection of the models specified in our pre-registration revealed that residuals were significantly non-normally distributed and had non-constant variance over groups. Since our data did not meet the assumptions for a linear modelling analysis, the analyses we present here instead evaluate our pre-registered predictions using 95% bootstrapped confidence intervals around the mean of each condition, as well as around the differences between condition means (following [12]). For each analysis, we used R [16] to calculate bias-corrected and accelerated (BCa) bootstrap intervals (as recommended by [15]) from 10,000 samples. The pattern of results from this analysis is identical to the one obtained from linear models.

[3]We tested for this interaction by generating, for each variation type, a set of bootstrap estimates of the difference between PRODUCTION LOAD and other memory load conditions (collapsed). The 95% bootstrapped confidence interval around the difference between these differences crosses zero, indicating no interaction.

above zero in all but the latter. On this measure, our data therefore suggest that there is a general bias towards introducing or boosting lexical conditioning, not arising from memory limitations during *production*. In fact, pairwise comparisons between conditions indicate that mutual information increased more in the UNPREDICTABLE/LEARNING LOAD condition than all others, suggesting that the preference for lexically-conditioned patterns of variation is actually amplified by memory limitations during *learning*.

To summarise, our results suggest that reduction in overall variation is driven by memory limitations during *production*. By contrast, a bias for lexical conditioning is found almost across the board, and is, if anything, more pronounced when memory is taxed during *learning* of unpredictable variation.

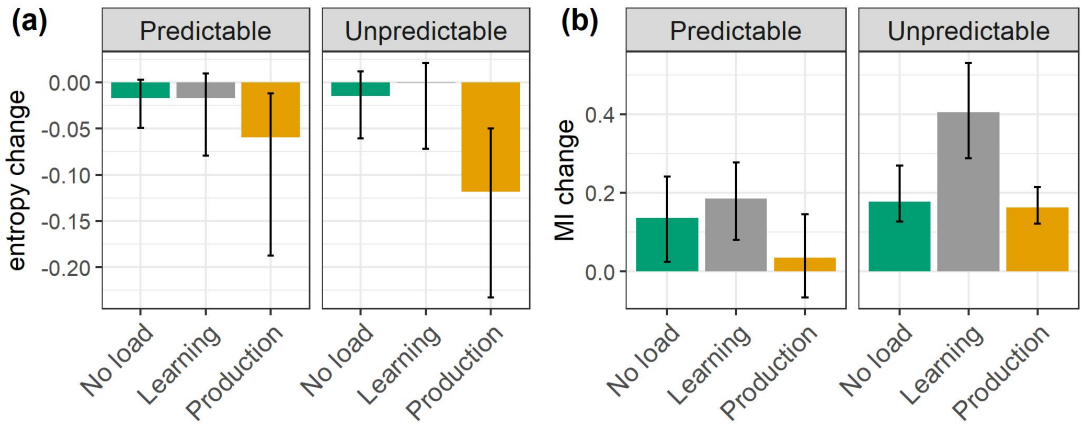


Figure 3: Change in entropy (left) and mutual information (right) from the training data to participants' productions, by condition. Error bars represent bootstrapped 95% confidence intervals over the mean.

4 Discussion

In this experiment, we investigated whether similar cognitive mechanisms drive regularisation of predictable (conditioned) and unpredictable (random) variation. Both types of variation exist in language, but there is historical evidence that both are lost over time, and experimental evidence that at least unpredictable variation is regularised – potentially due to memory limitations during production. Our findings partially align with previous research (e.g. [6, 11]) suggesting that regularisation is not driven by learners' failure to accurately encode the total amount of variation in their input. In particular, we found evidence for a reduction in overall variation during production, but not learning. This effect was indeed driven by memory load. However, our results point to a clear bias in favour of predictable patterns of variation, as evidenced by the increase in lexical conditioning during both learning and production. This effect was further amplified by memory limitations operating during learning, but is evident even in the absence of any memory load. In other words, a bias to reduce variability by increasing conditioning affects both language users' inferences during learning and their (implicit) decisions during production.

We began with the question of whether predictable and unpredictable variation are two points on the same spectrum. While our results suggest that some of the same mechanisms target both, we also find evidence that the introduction of lexical conditioning is not simply a

step on the way to eliminating variation altogether: the latter process appears to happen exclusively during production, while the increase in predictability is at least partially accounted for by learning effects. Furthermore, in both learning and production, we see much bigger changes in mutual information than in entropy, suggesting that the bias in favour of predictability is much stronger than the bias in favour of complete uniformity. A question for future work is how the relative strength of these biases interacts with the number of variants: with only two plural markers, as in our design, it is presumably not difficult to maintain both. Expanding the set of variants may heighten the pressure to reduce variation (although see [11]). A different kind of production task – with participants required to free-type their descriptions rather than being cued by having the whole lexicon presented as buttons – would also make retrieval more taxing [10]. We would predict that both of these modifications would result in greater entropy drop, especially under memory load during production.

It is also worth noting that, although our analysis shows a reliable reduction in entropy in the PRODUCTION LOAD conditions overall, there is considerable variation in how different participants respond to the memory manipulation. An open question is whether it is enough for only some individuals in a language community to have very strong biases against variation for that variation to be lost through cultural evolution. Certainly, the substantial individual differences we see here suggest at least that we should expect diachronic loss of variation from natural languages to be a slow process.

At a population level, though, our results suggest that language users have biases which work against variation of all kinds. From the perspective of language evolution, one might therefore wonder why variation of the kind described here is so pervasive in natural languages. As for any cognitive bias shaping language, the explanation for this is likely a combination of the fact that these biases are weak (i.e. defeasible), and compete with other pressures shaping language. Usage-based accounts (e.g. [1]) argue that systems of conditioned variation are maintained because the easiest variant to access in any given context is simply the one that has been experienced most often in that context (rather than the one that has been experienced most often overall). Furthermore, although this was not relevant in our task, this kind of variation may persist due to frequency-dependent patterns in regularity: irregular forms tend to be highly frequent, presumably making it easier to learn and retrieve the correct form [4].

5 Conclusion

We have provided evidence in favor of the claim that cognitive biases leading to regularisation target both unpredictable and predictable variation. Our findings support the idea that regularisation is particularly strong during production, and is driven at least in part by memory limitations. However, our results also suggest – contrary to our predictions and at least some previous work – that this memory bottleneck may play a distinct role during both learning and production. Specifically, while over-retrieval of the most accessible variant [8] during language production may act to reduce overall variability, *unpredictability* appears to decrease more as a result of inferences formed during learning. Overall, the results of this study lend support to the notion that cognitive constraints in individuals can give rise to particular structures in languages. By allowing languages to pass more easily through the bottleneck imposed by working memory limitations [2], we argue that properties like regularity are more likely to be selected through cultural evolution and thus develop into typological universals.

References

[1] Joan Bybee. From usage to grammar: The mind's response to repetition. *Language*, 82(4):711–733, 2006.

[2] Morten H. Christiansen and Nick Chater. The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, 2016.

[3] Jennifer Culbertson and Simon Kirby. Simplicity and specificity in language: Domain-general biases have domain-specific effects. *Frontiers in Psychology*, 6:1964, 2016.

[4] Christine Cuskley, Martina Pugliese, Claudio Castellano, Francesca Colaiori, Vittorio Loreto, and Francesca Tria. Internal and external dynamics in language: Evidence from verb regularity in a historical corpus of English. *PLOS ONE*, 9(8):e102882, 2014.

[5] Joshua R. de Leeuw. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1):1–12, 2015.

[6] Vanessa Ferdinand, Simon Kirby, and Kenny Smith. The cognitive roots of regularization in language. *Cognition*, 184:53–68, 2019.

[7] Talmy Givón. Function, Structure and Language Acquisition. In Dan Isaac Slobin, editor, *The Crosslinguistic Study of Language Acquisition*, pages 1005–1028. Lawrence Erlbaum Associates Inc., 1985.

[8] Adele Goldberg and Fernanda Ferreira. Good-enough language production. *Trends in Cognitive Sciences*, 26(4):300–311, 2022.

[9] Carla L. Hudson Kam. The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, 91:906–937, 2015.

[10] Carla L. Hudson Kam and Ann Chang. Investigating the cause of language regularization in adults: Memory constraints or learning effects? *Journal of Experimental Psychology: Learning Memory and Cognition*, 35(3):815–821, 2009.

[11] Carla L. Hudson Kam and Elissa L. Newport. Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, 59(1):30–66, 2009.

[12] Yasamin Motamedi, Lucie Wolters, Danielle Naegeli, Simon Kirby, and Marieke Schouwstra. From improvisation to learning: How naturalness and systematicity shape language evolution. *Cognition*, 228:105206, 2022.

[13] Andrew Perfors. When do memory limitations lead to regularization? An experimental and computational investigation. *Journal of Memory and Language*, 67(4):486–506, 2012.

[14] Andrew Perfors. Adult regularization of inconsistent input depends on pragmatic factors. *Language Learning and Development*, 12(2):138–155, 2016.

[15] Marie-Therese Puth, Markus Neuhäuser, and Graeme D. Ruxton. On the variety of methods for calculating confidence intervals by bootstrapping. *Journal of Animal Ecology*, 84(4):892–897, 2015.

[16] R Core Team. R: A Language and Environment for Statistical Computing, 2020.

[17] Florencia Reali and Thomas L. Griffiths. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, 111(3):317–328, 2009.

[18] Anna Samara, Kenny Smith, Helen Brown, and Elizabeth Wonnacott. Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, 94:85–114, 2017.

[19] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[20] Kenny Smith and Elizabeth Wonnacott. Eliminating unpredictable variation through iterated learning. *Cognition*, 116(3):444–449, 2010.

New evidence for the unlearnability of non-conservative quantifiers*

Tyler Knowlton[1], John Trueswell[2], and Anna Papafragou[3]

[1] MindCORE, University of Pennsylvania, Philadelphia, PA, U.S.A.
tzknowlt@upenn.edu
[2] Department of Psychology, University of Pennsylvania, Philadelphia, PA, U.S.A.
trueswel@psych.upenn.edu
[3] Department of Linguistics, University of Pennsylvania, Philadelphia, PA, U.S.A.
anna4@sas.upenn.edu

1 Introduction

The 'conservativity' of quantificational determiners is perhaps the most robust and renowned semantic universal [1, 7, 9, 21]. The typological generalization is this: duplicating a quantifier's internal (NP) argument in its external (predicative) argument is logically insignificant. Put another way, abstracting from variations in word order and morphology, the entailment pattern in (1) holds for any quantifier Q (e.g., *every fish swims* iff *every fish is a fish that swims*).

(1) [[Q NP] PRED] ↔ [[Q NP] [be NP that PRED]]

Potential exceptions to this pattern have been proposed. For example, *only*, if it were a determiner, would be a counterexample: *only fish swim* can be false when *only fish are fish that swim* is true. But in such cases, there are independent reasons for thinking the purported counterexamples are not quantificational determiners after all [6, 5, 2, 23, 16, 17].

That all quantifiers are conservative is significant, in part because it is unexpected given the standard view that quantifiers express relations between two sets. If we grant that languages have conservative quantifiers like *every* that express the subset relation in (2a), then what explains why languages lack a hypothetical non-conservative quantifier like *yreve*, which expresses the superset relation in (2b), or like *ident*, which expresses the identity relation in (2c)?

(2) a. ⟦Every fish swims⟧ ≈ FISH ⊆ SWIMMERS
b. ⟦Yreve fish swims⟧ ≈ FISH ⊇ SWIMMERS
c. ⟦Ident fish swims⟧ ≈ FISH = SWIMMERS

There is nothing conceptually more complicated about the latter two relations [20]. And it seems unlikely that an explanation can be found in terms of communicative need. After all, languages have expressions like *only* (*yreve* is *only* if it were a determiner instead of a focus operator) and *are identical to*. The fact to be explained is that no language seems to have a quantificational determiner with those sorts of meanings.

A reductionist (though logically possible) explanation would be that determiner conservativity is a historical accident. Maybe non-conservative meanings could be lexicalized as determiners, but they just haven't been in any natural language. Alternatively (and more interestingly), it has been argued that conservativity has an architectural explanation. It might be that only

*For helpful discussion, we thank Jeffrey Lidz, Paul Pietroski, Alexander Williams, and audiences at the UMass Amherst Psycholinguistics Workshop, BUCLD 47, and the University of Pennsylvania.

conservative relations and conservativity-preserving operations are part of the primitives out of which determiner meanings can be constructed [9], or that details of how syntactic movement is interpreted serve to filter out or otherwise disguise would-be non-conservative determiners [3, 13, 18], or that determiners never express non-conservative relations because they don't express relations in the first place [15, 11, 12, 14]. What these explanations have in common is that they all maintain that determiner conservativity reflects a fundamental fact about the language faculty, a constraint stemming from the architecture of the grammar, not one due to historical coincidence, communicative pressures, or domain general cognitive considerations. This claim predicts that children (and adults) should be unable to learn non-conservative quantifiers. But evidence bearing out this bold prediction has proven elusive.

2 Prior work on conservativity and learnability

Two studies have previously looked for evidence of a learnability advantage for novel conservative over novel non-conservative determiners. In one [8], 20 5-year-old children were introduced to a picky puppet who 'likes' certain scenes and 'dislikes' others. The child's task was to help the experimenter sort scenes into piles according to whether the puppet liked them. Children were told that the puppet likes it when "gleeb girls are on the beach", and were left to discern what *gleeb* means. In the Conservative condition, *gleeb* meant *not all*, so "gleeb girls are on the beach" was true just in case there was at least one girl not on the beach (i.e., there was one girl in the grass). In the Non-Conservative condition, *gleeb* meant *not only*, so "gleeb girls are on the beach" was true just in case there was at least one boy on the beach. After watching the experimenter sort five cards, children were presented with five new cards and asked to sort them. They showed the predicted pattern: participants in the Conservative condition performed significantly better than those in the Non-conservative condition, and only participants in the Conservative condition performed significantly above chance.

In an attempt to replicate this result, though, another study [19] found that children showed no evidence of learning in either condition. Children also failed to show the expected result when the task was modified to allow the puppet to 'correct' the situation to his liking (e.g., by moving a character). Moreover, the effect failed to replicate in 18 English-speaking adults.

The experiments reported below aim to improve on these previous designs in the hopes of demonstrating that non-conservative meanings cannot be learned as quantificational determiners even when (i) their conservative counterparts can be learned as quantifiers in the same experimental setting and (ii) they can be learned as instances of other syntactic categories.

3 Current experiments

The current experiments differ from the prior studies cited in Section 2 in a number of ways. First, instead of using *not all* and *not only* as the novel conservative and non-conservative meanings, the current experiments compare the conservative *all but one* against the non-conservative *outnumbers by one*. That is, *gleeb frogs are green* would mean "all but one frogs are green" in the Conservative condition and "the frogs outnumber the green things by one" in the Non-conservative condition. The rationale behind this change is that the novel meanings used above, *not all* and *not only*, are both negations of existing quantifier meanings, and learners may be slower to learn meanings that involve negation [4]. Since the failure to replicate discussed above is driven by a lack of learning in both conditions, trying a different pair of novel quantifiers seemed sensible.

Second, instead of being asked to sort scenes into piles based on whether a puppet liked them (which may have highlighted the puppet's preferences more than the novel word's meaning), participants here were explicitly told that their job is to learn a novel word. And their knowledge was tested in ways that served to highlight the word-learning aspect of the task (e.g., by being asked whether a sentence with *gleeb* is true relative to a picture).

Third, the experiments presented here use sentences in the partitive frame (e.g., *gleeb of the circles are blue*), which provide an unambiguous signal of the novel word being a determiner. In contrast, past work avoided the partitive, opting instead for sentences like *gleeb girls are on the beach*. This is a sensible choice: the partitive frame might encourage attention to the quantifier's internal argument as such, leading participants to focus on the girls or the circles independent of their hypothesized meaning for *gleeb*. But the lack of partitive may have led some participants to posit a different syntactic category for *gleeb* (e.g., *gleeb* could be a focus operator like *only* or an adjective like *friendly*). And while phrases like *of the circle* do likely encourage attending to the circles, quantificational phrases that lack the partitive, like *every circle*, have since been shown to drive attention to the circles as well [10, 11]. This fact and conservativity likely share a common cause, making it difficult and perhaps futile to control for.

Finally, the current study departs from previous work in that it focuses only on adult participants. This is in part out of convenience, but also because adults have the full complement of cognitive resources at their disposal. If they are able to learn novel conservative quantifiers but unable to learn novel non-conservative ones, it is hard to imagine children faring any better. Of course, it remains a possibility that adult participants can fail to learn in such a task by 'overthinking' in a way that child participants will not. So there is certainly value in adapting the experiments presented below for children. But one would like to know if the effect can robustly be observed in adults first, so we restrict our attention to adult participants here.

Each experiment is discussed in its own subsection below. But for ease of display, all data are plotted together in Figure 1, along with example training and test trials for each experiment. The prediction throughout – on the hypothesis that conservativity has a grammatical explanation – is that only participants in the Conservative condition will successfully learn the meaning of a novel quantifier *gleeb*, but that any such learnability advantage will disappear if *gleeb* is instead taught as a novel verb.

3.1 Experiment 1: Learning by example

In Experiment 1, participants (60 English-speaking adults recruited on `prolific.co`) were trained with a series of example images and corresponding sentences using the novel quantifier *gleeb*. Participants in the Conservative condition were given training trials consistent with the sentence in (3) having the meaning in (3a), whereas those in the Non-conservative condition were given training trials consistent with (3) having the meaning in (3b).

(3) Gleeb of the circles are blue.

a. $|\textsc{circles}| - 1 = |\textsc{blue-circles}|$ *all but one of the circles are blue*

b. $|\textsc{circles}| - 1 = |\textsc{blue-things}|$ *the circles outnumber the blue things by one*

Training consisted of 16 images of blue and orange circles and squares. Each picture was either described with the sentence "gleeb of the circles or blue" or the sentence "it's not the case that gleeb of the circles are blue", depending on the condition. Participants were also explicitly told the number of circles and blue shapes present in each display (this information is more helpful in the Non-conservative condition and was included to tilt the odds in favor of success in that condition). All text presented on screen was also read aloud to participants using a speech

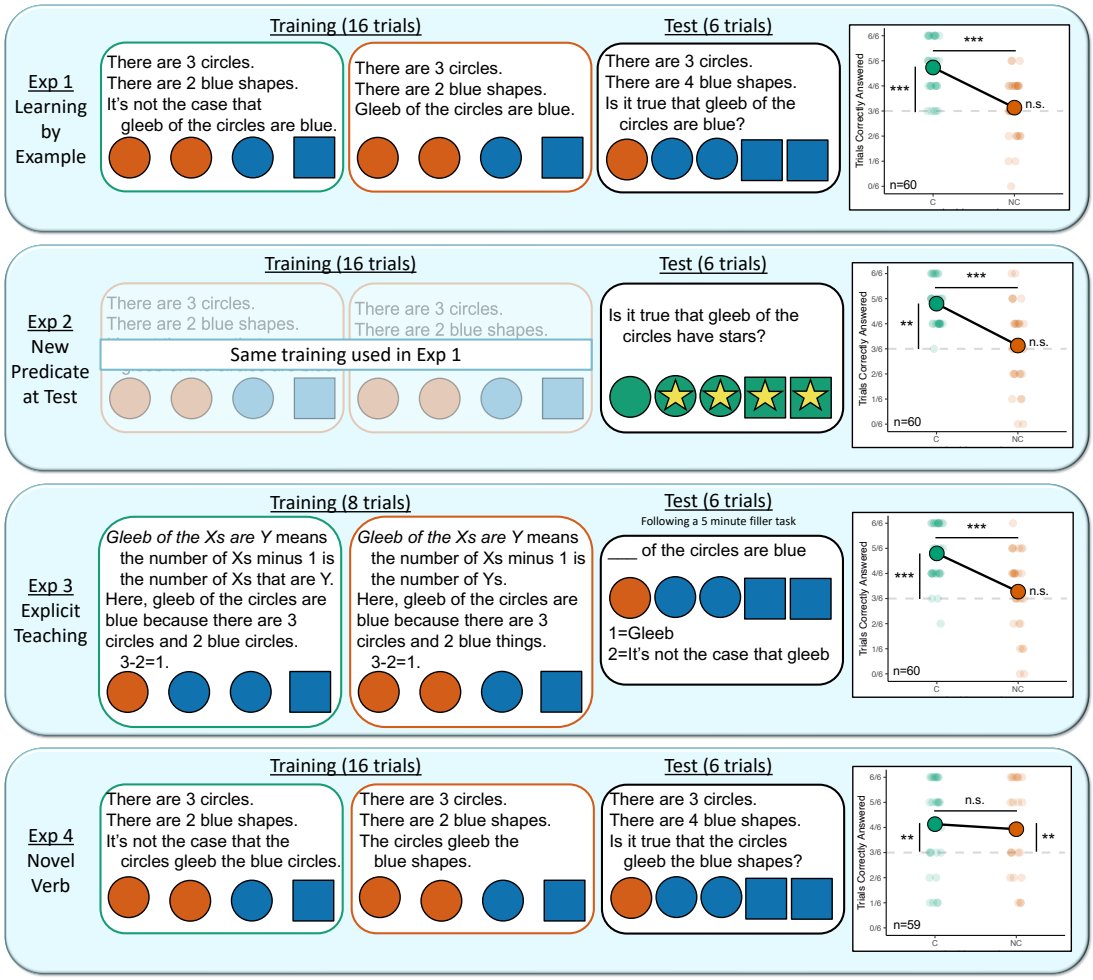


Figure 1: Example stimuli and results from all four experiments. Starting from the left, example training trials from the Conservative (green outline) and Non-conservative (orange outline) conditions are shown. Then, one example test trial (black outline) is shown. Results for each experiment are shown on the far right. Large points represent average performance, translucent points represent the number of test trials each individual participant correctly answered. Chance performance, 3/6, is represented by the grey dashed line. Stars represent significance of mixed-effects model intercepts and coefficients. Experiments were designed in PCIbex [22].

synthesizer. The particular training trials used were designed to rule out easily-hypothesized meanings for *gleeb* (e.g., training trials were included where *more than half of the circles are blue* was false but *gleeb of the circles are blue* was true).

Immediately after training, participants were given six new images and asked "is it true that gleeb of the circles are blue?". For three images, the correct answer was yes; for the other three images, the correct answer was no. Both the training and the test images were held constant across conditions, but the correct answer flipped (a true/'yes' image in the Conservative

condition would be a false/'no' image in the Non-conservative condition). Put another way, participants in both conditions saw the exact same images; what differed was the sentences paired with those images.

Results. Participants in the Conservative condition were more accurate than those in the Non-conservative condition (comparing a model with condition as a fixed effect to an intercept-only model with the same random effects structure: $\chi^2(1) = 23.28, p < .001$; main effect of condition: $\beta = .66$ [95% CI .52 to .79], $z = 4.93, p < .001$). And while those in the Conservative condition performed significantly above chance (intercept: $\beta = 1.72$ [95% CI 1.22 to 2.22], $z = 3.43, p < .001$), those in the Non-conservative condition did not (intercept: $\beta = .09$ [95% CI $-.06$ to .24], $z = 0.59, p = .553$). That is, only participants in the Conservative condition showed evidence of learning the novel meaning, as predicted.

3.2 Experiment 2: New predicate at test

In Experiment 2, participants (60 English-speaking adults) were trained exactly as in Experiment 1 (with 16 pictures of blue and orange circles and squares). The only difference came in the test phase: instead of differing in color, the shapes were all green, and they either had or lacked yellow stars. Everything else about the distribution of shapes was the same as in Experiment 1: the shapes that were blue in Experiment 1 had stars in Experiment 2, and those that were orange in Experiment 1 lacked stars in Experiment 2. Importantly, the test sentences used a different predicate ("is it true that gleeb of the circles have stars?"). This change helps ensure that participants who succeed at the task actually do so by learning the meaning of *gleeb*, as opposed to relying on visual similarities between the training and test images.

Results. The results from Experiment 1 were replicated. Participants in the Conservative condition were more accurate than those in the Non-conservative condition ($\chi^2(1) = 21.03, p < .001$; main effect of condition: $\beta = .73$ [95% CI .58 to .88], $z = 4.75, p < .001$). Moreover, while those in the Conservative condition performed significantly above chance (intercept: $\beta = 2.04$ [95% CI 1.33 to 2.75], $z = 2.87, p < .01$), those in the Non-conservative condition did not (intercept: $\beta = .1$ [95% CI $-.12$ to .32], $z = 0.46, p = .647$). As in Experiment 1, then, only participants in the Conservative condition showed evidence of learning the novel meaning.

3.3 Experiment 3: Explicit teaching

Given the difficulties experienced in the Non-conservative condition in Experiments 1 and 2, Experiment 3 sought to make the task easier through an 'explicit teaching' paradigm. In this version, participants (60 English-speaking adults) were explicitly told the meaning of *gleeb* before being guided through the training trials (e.g., in the Non-conservative condition: "*Gleeb of the Xs are Y* means that the number of Xs minus one is the number of Ys"). Additionally, an explanation was given as to why each image made the sentence true or false (e.g., "Here, gleeb of the circles are blue because there are 3 circles and 2 blue shapes and 3-2=1").

After completing eight training trials, participants were given a 5-minute unrelated filler task before moving on to the test phase. Here, the test phase asked them to fill in a blank in a sentence like "___ of the circles are blue" (with either "Gleeb" or "It's not the case that gleeb"). This was meant to remind them that they learned a novel word in the training portion (and to discourage strategies based on pattern matching).

Results. Even with the 'brute force' nature of the training, Experiment 3 again replicates the effect found in Experiments 1 and 2. Participants in the Conservative condition were more accurate than those in the Non-conservative condition ($\chi^2(1) = 16.34, p < .001$; main effect of condition: $\beta = .69$ [95% CI .52 to .85], $z = 4.05, p < .001$). And while those in the Conservative condition performed significantly above chance (intercept: $\beta = 1.74$ [95% CI 1.27 to 2.21], $z = 3.72, p < .001$), those in the Non-conservative condition did not (intercept: $\beta = .21$ [95% CI $-.01$ to .43], $z = 0.94, p = .346$). So despite being explicitly taught in both conditions, only those in the Conservative condition were able to successfully encode and remember *gleeb*'s meaning.

3.4 Experiment 4: Novel verb instead of novel determiner

The results of Experiments 1-3 suggest that pairing a non-conservative meaning with a novel determiner is at least exceedingly unnatural, if not outright impossible. But since conservativity is specific to quantificational determiners, any learnability asymmetry that exists should disappear if the novel word is of a different syntactic category. Experiment 4 aimed to test this prediction; to make the effect observed in Experiments 1-3 go away. To do so, participants (59 English-speaking adults) were taught the same two novel meanings, but *gleeb* was introduced as a verb instead of as a determiner. In the Conservative condition, images were described with sentences like "The circles gleeb the blue circles". In the Non-conservative condition, images were described with sentences like "The circles gleeb the blue shapes." Otherwise, the same scene-sentence pairs were used as in the above three experiments.

Results. As predicted, the learnability advantage for the conservative *gleeb* disappeared when both novel words were taught as verbs. Participants in the Conservative condition were not significantly more accurate than those in the Non-conservative condition ($\chi^2(1) = 0.27, p = .601$; no effect of condition: $\beta = .11$ [95% CI $-.1$ to .33], $z = 0.53, p = .598$). And both groups performed significantly better than chance (Conservative intercept: $\beta = 1.13$ [95% CI 0.78 to 1.48], $z = 3.22, p < .01$; Non-conservative intercept: $\beta = 0.8$ [95% CI 0.53 to 1.08], $z = 2.93, p < .01$). This suggests the results from Experiments 1-3 were driven by leaning the non-conservative meaning as a quantifier per se, not any difficulties with the meaning itself.

4 Conclusion

Cross-linguistically, all quantifiers are conservative. The robustness of this semantic universal invites a learnability claim: if conservativity reflects a deep fact about the language faculty, then non-conservative quantifiers should be impossible to learn. The four experiments presented above bear out this prediction. Adults were able to pair a novel conservative meaning, but not its non-conservative counterpart, with a novel quantifier. And, as predicted, this effect disappeared when the novel meanings were instead taught as novel verbs. In future work, we plan to confirm that this finding generalizes beyond the particular novel meanings tested here (*all but one* vs. *outnumbers by one*) to other pairs of conservative and non-conservative meanings. Moreover, we intend to extend the current task to children. But in the meantime, the experiments reported above lend support to views on which conservativity is a cornerstone of the semantics of determiners (e.g., [11, 15, 14]).

References

[1] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. In *Philosophy, Language, and Artificial Intelligence*, pages 241–301. Springer, 1981. https://www.jstor.org/stable/25001052.

[2] Ariel Cohen. Relative readings of many, often, and generics. *Natural Language Semantics*, 9:41–67, 2001. http://dx.doi.org/10.1023/A:1017913406219.

[3] Danny Fox. Antecedentcontained deletion and the copy theory of movement. *Linguistic Inquiry*, 33(1):63–96, 2002. https://doi.org/10.1162/002438902317382189.

[4] Angela Xiaoxue He and Alexis Wellwood. "Most" is easy but "least" is hard: Novel determiner learning in 4-year-olds. In Jennifer Culbertson, Andrew Perfors, Hugh Rabagliati, and Veronica Ramenzoni, editors, *Proceedings of the 44th Annual Conference of the Cognitive Science Society*, 2022. https://escholarship.org/uc/item/5hh4m526.

[5] Elena Herburger. Focus and weak noun phrases. *Natural Language Semantics*, 5(53):53–78, 1997. https://doi.org/10.1023/A:1008222204053.

[6] Elena Herburger. *What counts: Focus and quantification*. MIT Press, 2000.

[7] James Higginbotham and Robert May. Questions, quantifiers and crossing. *The Linguistic Review*, 1(1):41–80, 1981. http://dx.doi.org/10.1515/tlir.1981.1.1.41.

[8] Tim Hunter and Jeffrey Lidz. Conservativity and learnability of determiners. *Journal of Semantics*, 30(3):315–334, 2013. doi.org/10.1093/jos/ffs014.

[9] Edward L Keenan and Jonathan Stavi. A semantic characterization of natural language determiners. *Linguistics and Philosophy*, 9(3):253–326, 1986. https://www.jstor.org/stable/25001246.

[10] Tyler Knowlton, Paul Pietroski, Justin Halberda, and Jeffrey Lidz. The mental representation of universal quantifiers. *Linguistics and Philosophy*, 45:911–941, 2022. https://doi.org/10.1007/s10988-021-09337-8.

[11] Tyler Knowlton, Paul Pietroski, Alexander Williams, Justin Halberda, and Jeffrey Lidz. Determiners are "conservative" because their meanings are not relations: Evidence from verification. *Semantics and Linguistic Theory*, 30:206–226, 2021. https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/view/30.206.

[12] Peter Lasersohn. Common nouns as modally non-rigid restricted variables. *Linguistics and Philosophy*, 44(2):363–424, 2021. https://doi.org/10.1007/s10988-019-09293-4.

[13] Peter Ludlow. LF and natural logic. In Gerhard Preyer and Georg Peter, editors, *Logical form and language*, pages 132–168. Oxford University Press, 2002.

[14] Peter Ludlow and Sašo Živanović. *Language, Form, and Logic: In Pursuit of Natural Logic's Holy Grail*. Oxford University Press, 2022. https://doi.org/10.1093/oso/9780199591534.001.0001.

[15] Paul Pietroski. *Conjoining Meanings: Semantics Without Truth Values*. Oxford University Press, 2018. https://doi.org/10.1093/oso/9780198812722.001.0001.

[16] Maribel Romero. The conservativity of *many*. In *20th Amsterdam Colloquium*, pages 20–29, 2015. https://ling.sprachwiss.uni-konstanz.de/pages/home/romero/papers/romero_v8-AC20.pdf.

[17] Maribel Romero. The conservativity of *many*: Split scope and *most*. *Topoi*, 37:393–404, 2018. https://doi.org/10.1007/s11245-017-9477-5.

[18] Jacopo Romoli. A structural account of conservativity. *Semantics-Syntax Interface*, 2(1):28–57, 2015. https://pure.ulster.ac.uk/en/publications/a-structural-account-of-conservativity-3.

[19] Jennifer Spenader and Jill de Villiers. Are conservative quantifiers easier to learn? evidence from novel quantifier experiments. In *22nd Amsterdam Colloquium*, 2019. http://hdl.handle.net/11370/79051821-bdba-4168-8214-5cf1033b8451.

[20] Shane Steinert-Threlkeld and Jakub Szymanik. Learnability and semantic universals. *Semantics*

and Pragmatics, 12(4), 2019. https://doi.org/10.3765/sp.12.4.

[21] Kai von Fintel and Lisa Matthewson. Universals in semantics. *The Linguistic Review*, 25(1-2):139–201, 2008. https://doi.org/10.1515/TLIR.2008.004.

[22] Jeremy Zehr and Florian Schwarz. PennController for internet based experiments (IBEX). 2018. https://doi.org/10.17605/OSF.IO/MD832.

[23] Richard Zuber. A class of non-conservative determiners in polish. *Lingvisticæ Investigationes*, 27(1):147–165, 2004. https://doi.org/10.1075/li.27.1.07zub.

Generalizing Polarity Sensitive Items to Positive Environments in an Artificial Language

Jeremy Kuhn[1] and Mora Maldonado[2]

[1] Institut Jean Nicod, CNRS, ENS, EHESS, PSL University
jeremy.kuhn@ens.psl.eu
[2] LLING, CNRS, Université de Nantes
mora.maldonado@univ-nantes.fr

Abstract

Two well known patterns involve polarity sensitive items. First is Jespersen's cycle: a polarity sensitive item appearing in the scope of negation is reanalyzed as itself the marker of negation. Second, many NPIs in English have a secondary positive use, with a systematic semantic relation to the polarity sensitive use: in each case, the meaning of the positive form is the logical dual of the NPI. This study investigates these meaning changes using an artificial language learning paradigm, where learners are trained on ambiguous input, and must generalize beyond this input. We ask which generalization strategies appear when learners have to interpret polarity sensitive items in a positive environment.

1 Introduction

Two well known patterns involve polarity sensitive items. The first is a diachronic pattern known as Jespersen's cycle [3, 5]: a polarity sensitive item appearing in the scope of negation is often (historically) reanalyzed as itself the marker of negation, as shown for French in (1).

(1) Jespersen's cycle
a. Je ne comprends. → b. Je ne comprends pas. → c. Je comprends pas.

The second pattern regards Negative Polarity Iitems (NPIs), which often have a secondary positive use [6], as shown for English in (2). Ladusaw observes that the semantic relation between the two uses is systematic: in each case, the meaning of the positive form is the logical dual, the meaning one would have to assign if the NPI were scoping over negation, as illustrated in (3). Other languages have similar patterns, e.g. French *encore*, 'yet/still'.

(2) Dual uses of NPIs
a. I talked to any person who'd listen. (*any* = *every*)
b. I've been here ever since you called. (*ever* = *always*)
c. The winner is yet to be decided. (*yet* = *still*)
d. Gas is expensive anymore. (*anymore* = *nowadays*)

(3) a. not any = every not
b. not ever = always not
c. not yet = still not
d. not anymore = nowadays not

1.1 A diachronic view

These meaning shifts can be understood from a diachronic perspective. Let us start by considering a child learning a language with polarity sensitive items. In principle, the child should be

able to correctly identify the meaning of the sentence as a whole. However, due to distributional constraints (i.e., the fact that some items consistently co-occur together), multiple equally viable analyses are possible for the meaning of the individual words and the way they are put together. If the syntactic constraints are relaxed, as it may happen in language evolution, these different analyses would lead to observably different sentential meanings.

For Jespersen's cycle, consider a snapshot of a language at stage (1)b. A child learning language will successfully understand that the sentence as a whole has a negative meaning (in this case, 'the speaker doesn't understand'), and can deduce the meanings of *je* and *comprends* from their use in other sentences. But, because *ne* and *pas* always occur together, the child does not know which element contributes the negation—there are two unresolved variables. Postulating that *pas* has negative meaning allows the development of stage (1)c, where the syntactic constraint of co-occurrence has been removed.

Dual uses of NPIs can be viewed through a similar diachronic lens. Consider a snapshot of a language in which a particular lexical item always occurs in the presence of negation. In such a situation, two equally viable analyses are possible to derive the meaning of a sentence containing the item: either it has one meaning (e.g., $\exists$) and scopes below negation, or it has another meaning (e.g., $\forall$) and scopes above negation. Once again, there are two unresolved variables: the denotation of the lexical item, and its scope (above or below negation).

If the syntactic constraint is removed (e.g., if an NPI appears in a non-negative context), this should reveal the (preferred) underlying interpretation of the polarity sensitive item as well as its relative scope. The examples in (2) thus suggest a stage where many NPIs are interpreted as wide-scoping elements with the appropriate (dual) meaning. Note that this is not incompatible with the existence of another (earlier or later) stage with the same items are interpreted as narrow-scoping.

For both processes described above, when a particular lexical item is restricted to a specific syntactic environment, there may be a systematic ambiguity of analysis that, in its resolution, ends up acting as a vector of grammatical change.

1.2 Current study

The present study investigates these possible meaning changes using an artificial language learning paradigm, where learners are trained on ambiguous input, and must generalize beyond this input in a way that is compatible with only one of the hypotheses of interest (a.k.a. 'Poverty of stimuli' paradigm [2, *i.a.*]). Here, we ask which generalization strategies appear when learners have to interpret polarity sensitive items in a positive environment.

For sentences of the form [... Neg ... X ...], where X only appears in the presence of a word independently learned to denote sentential negation, three possible generalization strategies are possible when X appears on its own:

(4) (i) a narrow-scope (often existential) meaning (=the 'literal' NPI meaning);
(ii) a wide-scope (often universal) meaning (=its logical dual); or
(iii) a redundant negative meaning, analogous to Neg itself (=Jespersen's cycle).

In Experiment 1 (Exp 1), English-speaking participants were taught an artificial language that included a negative marker and a polarity sensitive degree modifier roughly equivalent to English *at all*. All participants were trained on sentences where this item always co-occurred with negation. That is, the input did not include any example of this item in absence of negation. At test, learners were asked to assign an interpretation to sentences where this novel item appeared on its own, without negation. In Experiment 1a (Exp 1a), we contrasted strategies (i)/(iii) and (ii); in Experiment 1b (Exp 1b), strategies (i)/(ii) and (iii). Experiment

2 (Exp 2) repeated Exp 1a but with a polarity sensitive item quantifying over times, akin to English *ever*.

All experimental materials, data, and analysis reported here are available at osf.io/kzju8. Experimental design and analysis were preregistred for Exp1a (Exp 1b is identical) and Exp2.

2 Experiment 1

2.1 Methods

Design and Materials The miniature language consisted of four predicates, four proper nouns, one negative marker (*em*) and one degree modifier (*tup*), akin to English *at all*. All words were created following English phonotactics and presented orthographically. Predicates denote gradable properties with closed scales: transparent/opaque, open/closed, top/bottom, and standing/lying. For each property, we define four possible scale points (or degrees): 'minimum', 'near-minimum', 'near-maximum', and 'maximum', as shown in Table 1 for 'transparent'. The predicate-meaning mapping was manipulated between subjects such that the maximum scale point could be linked to either end of the scale (e.g., the maximum degree could be 'opaque' or 'transparent').

Participants were exposed to non-target and to critical sentences. Non-target sentences were: *simple positive (SP)* (e.g., *Greenie pleet*), *simple negative (SN)* (e.g., *Greenie em pleet*), and *negative 'at all' (Neg-'At all')* (e.g., *Greenie em pleet tup*). The meaning of these sentences was conveyed by showing participants whether they could be used to describe scenarios where the predicate applied to the different degrees, as in Table 2a.

The critical configuration, seen only during testing, was the *positive 'at all' (Pos-'At all')* sentence, where the modifier *tup* appeared on its own, without negation (e.g., *Greenie pleet tup*). In Exp 1a, participants had to decide whether Pos-'At all' sentences described a scenario where the predicate applied to a maximal or to a near-minimal degree. In Exp 1b, participants were asked to choose whether the predicate applied to a maximal or to a minimal degree.

Exp 1a and 1b together allow us to dissociate the generalization patterns in (4): If learners interpret *tup* as an universal quantifier, they are expected to go for the 'maximum degree' interpretation across the board. Similarly, if *tup* is treated as a redundant negative marker, both 'near-minimum' and 'minimum degree' interpretations would be compatible with the sentence. In contrast, if *tup* is treated as an existential quantifier, learners are expected to go for the 'near-minimum degree' reading in Exp 1a but for the 'maximum degree' one in Exp 1b.

Procedure The experimental procedure was the same in Exps 1a and 1b. Participants were instructed that they were going to learn a foreign language called Hiri Motu through a training and a testing phase. In the training phase, participants were taught how to interpret non-target sentences. The training consisted of (a) an exposure block, where participants were presented with scenarios that verify the sentence, and (b) a sentence-picture matching block, where participants were presented with a sentence together with two pictures and had to select which picture best described the sentence (see Fig. 1A). Participants were given feedback on their answers. There were 14 exposure trials and 24 sentence-picture matching trials.

The testing phase involved a similar matching task but also included trials for the critical, held-out Pos-'At all' sentence. The alternative pictures for this held-out sentence were determined by the experiment (see Fig. 1B). There were 32 testing trials (8 involving the critical sentence configuration). No feedback was provided to participants at this stage.

At the end of the experiment, participants were asked to translate to English four sentences from the artificial language, one of each type.

Participants 132 English-speaking adult participants were recruited via Prolific (Exp 1a: 75, Exp 1b: 57). Participants were paid 9£/hour for their participation which lasted on average 15 minutes. Per our pre-registration, we only included in the analysis participants who were successfully trained on non-target sentences (i.e. accuracy rates above 75%). This resulted on the analysis of 49 participants in Exp 1a and 33 in Exp 1b.

2.2 Results and discussion

Fig. 3a shows the proportion of trials on which participants chose the 'maximum degree' scenario for Pos-'At all' sentences during the test phase. A visual inspection suggests that in both experiments about a quarter of participants interpreted Pos-'At all' sentences as compatible with scenarios where the predicate applies to maximum degree, but half the participants generalized to the other choice, whatever it was. Per our pre-registration, we evaluated this pattern statistically with logistic mixed-effects regression models, including random intercepts per subject.[1] These models revealed that the proportion of 'maximum-degree' responses was significantly below chance in Exp 1a ($\beta = -1.563; p = .0137$) and not different from chance in Exp 1b ($\beta = -0.89; p = .2$).[2]

In order to compare the two experiments, we fit a third logistic regression, predicting responses by Experiment. The effect of Experiment was not significant ($\beta = .69$; $p = .46$). This suggests that participants who interpreted Pos-'At all' sentences as compatible with 'near-minimum degree' scenarios in Exp 1a learned something analogous to a redundant negative meaning, and not a proper existential meaning.

Translations provided at the end of Exp 1a give further support for this interpretation. 'Maximum degree' responders systematically translated Pos-'At all' sentences with universal degree modifiers like *completely* or *very* and Neg-'At all' sentences as involving either the same words (e.g. *Greenie is very transparent*) or the modifier *at all* (e.g. *Greenie is not transparent at all*). Among the 'near-minimum degree' responders, translations of Pos-'At all' sentences were less consistent; no subject translated the meaning using existential degree modifiers like *a bit*, but a number of participants gave the sentence a *negative* meaning.

Altogether, Exp 1 results suggest that learners tend to interpret the degree modifier *tup* as an additional, redundant negation. This modifier can co-appear with negation, as in Neg-'At all' sentences, giving rise to an emphatic negative meaning, or appear on its own, leading to the same meaning as the alternative negative marker *em*.

Now, polarity sensitivity is known to vary within and across languages, possibly as a function of the lexical semantics of the items themselves [4]. In Exp 1, 'at all' was chosen as the target meaning because English *at all* is one NPI that that does not have a positive dual use. On the other hand, it is possible that lexical semantic properties of *at all* also allow it to be more easily interpreted as a bleached marker of sentential negation. In Exp 2, we thus modified the experimental materials so the artificial language involves a different polarity-sensitive item: a temporal adverb roughly equivalent to the English word *ever*.

[1] All statistical analyses were performed using the lme4 package in R [7, 1]. In the models reported here, random intercepts per predicate and nesting by group had to be removed due to lack of convergence.

[2] The low number of participants in Exp 1b might have resulted on a lack of sensitivity to the effect. In an initial stage, the exclusion threshold for Exp 1b was of 66% accuracy rates in non-target trials (rather than 75%); only later we noticed that this was not in accordance with what was preregistered for Exp 1a. Analysis of the data with the lower exclusion threshold showed a statistical effect in the same direction as Exp 1a.

3 Experiment 2

3.1 Methods and materials

The methodology in Exp 2 was analogous to Exp 1a, except in the details specified below.

Design and Materials Participants were presented with a miniature language which in this case included one copula, three predicates, one proper noun, one negative marker (*mo*) and one adverb (*yele*), akin to English *ever*. Predicates denote three possible locations in a map: city, desert, forest. All sentences in the artificial language refer to either the current location or the past trajectory of a character, Jojo. For example, a sentence can express that Jojo is currently in the city, or that they have never been in the city. For each location, we define four possible temporal states of Jojo: 'never', 'previously', 'currently', and 'always', as shown in Table 3 for the city location. Note that one can establish a correspondence between Jojo's temporal states for each location, and the scale points or degrees used in Exps 1a and 1b.

As in Exp 1, participants were first trained on three non-target sentences: (i) *simple positive (SP)* (e.g., *Jojo es villum*), used either when Jojo has always been or currently is in the location denoted by the predicate; (i) *simple negative (SN)* (e.g., *Jojo mo es villum*), used either when Jojo has previously been or has never been in the location denoted by the predicate; and (iii) *negative 'Ever' (Not-'Ever')* (e.g., *Jojo mo es villum yele*), used exclusively when Jojo has never been in the location denoted by the predicate.

At test, participants were asked to additionally provide a meaning for *positive 'ever' (Pos-'Ever')* sentences (e.g., *Jojo es villum yele*), where *yele* does not co-appear with negation. Participants had to decide whether this critical sentence means that Jojo has always been in the given location or whether they have previously been in this location (but no longer are). As in Exp 1a, this choice serves to distinguish between strategies (i)/(iii) vs. (ii) in (4).

Procedure The procedure was analogous to Exp 1. There was a training and a testing phase, including 22 exposure trials, 32 training trials and 32 testing trials (8 involving the Pos-'Ever' sentences). A trial example is provided in Fig. 2.

Participants We recruited 59 English-speaking participants via Prolific. 31 participants were successfully trained on non-target sentences and taken into account for the analysis.[3]

3.2 Results and discussion

Fig. 3b shows the proportion of trials on which participants chose the 'always' scenario for Pos-'Ever' sentences during testing. Per our pre-registration, we fit a logistic mixed effects regression (random intercepts per subject) to statistically evaluate these responses. Our model reveals that the log-odds of selecting the 'always' scenario is significantly above chance ($\beta = 9.5; p < .001$). This result contrast with Exp 1, as a majority of participants appear to consistently take the polarity item to have a universal meaning in positive environments, consistent with a logical dual interpretation (i.e., strategy (ii) in (4)).

4 General Discussion

Our findings reveal that different strategies are adopted when extending the meaning of polarity sensitive items to contexts without a licensor. These strategies correspond to two meaning shifts

[3]Training items turned out to be difficult to learn, and for this reason we have not been able to collect our target sample size in this experiment.

attested cross-linguistically: (i) Jespersen's cycle and (ii) positive wide-scope duals. In Exp 1a, most participants assign a 'near-minimum degree' interpretation to Pos-'At all' sentences. The comparison with Exp 1b suggests that this is not due to an existential interpretation, but rather to a negative interpretation of the NPI, possibly due to a repair strategy with reconstructed negation. This generalization corresponds to Jespersen's cycle, in which a minimizing NPI is reinterpreted as contributing negation itself.

In Exps 1a and 1b, we additionally observe a small group of participants who assign a 'maximum degree' interpretation to Pos-'At all' sentences, thus displaying a pattern of generalizing the NPI meaning to its wide-scope dual. Exp 2 corroborates the existence of this group, and suggests that strategies differ depending on the semantic properties of the polarity sensitive item in question: *at all* is a better marker of redundant negation than *ever*.

Differences between NPIs Exps 1 and 2 displayed different generalization preferences. Since the two tasks were not perfectly parallel, it is possible that these differences arose in part due to irrelevant aspects of the experimental design. On the other hand, it is also possible the different generalization preferences arise from differences in the lexical semantics of the two NPIs. Why might *at all* be a better marker of redundant negation than *ever*?

One possible explanation concerns the relation of the NPIs to a certain kind of context-sensitivity. Specifically, simple negative sentences show vagueness with respect to degree quantification that they do not show with respect to time quantification. Consider the sentence in (5) with the two scenarios described in (6). While the sentence could easily be judged to be true in either scenario, one can nevertheless judge Scenario (b) as a 'better' instantiation of the sentence. In contrast, consider the same sentence with the two scenarios in (7). Again, the sentence is true in either scenario. Here, though, there is no clear intuition that either scenario is 'better' than the other.

(5) The gas tank is not full.

(6) a. The tank is a quarter full.
b. The tank is completely empty.

(7) a. The tank is completely empty, but was full yesterday.
b. The tank is completely empty, and has been all week.

The difference in intuitions can be explained in terms of context-sensitivity. Notably, (5) has a context-sensitive degree variable that may take different values depending on the question under discussion. (One needs to know *how full* counts as 'full' to evaluate the sentence.) In contrast, the time variable in (5) shows no context-sensitivity, as it is saturated by the time of utterance. (One doesn't need to know *how often* the tank is full to evaluate the sentence.)

As a consequence, the meaning of *at all* can more easily be bleached than the meaning of *ever*. In certain contexts, the threshhold of *full* is sufficiently low that (5) is synonymous with *The tank is not at all full*. In contrast, *ever* introduces temporal quantification not present in (5); there is thus no context in which (5) is synonymous with *The tank is not ever full*. The inability to ignore the temporal component of *ever* may thus explain why it resists the bleached negative meaning that appears as the dominant generalization strategy in Experiment 1.

The influence of context-sensitivity, as well as other aspects of the lexical semantics, can be investigated in follow-up experimental work with other polarity sensitive items ('yet', 'anymore'), as well as typological work on sycretisms involving NPIs cross-linguistically.

References

[1] Douglas M Bates. lme4: Mixed-effects modeling with r, 2010.

[2] Jennifer Culbertson and David Adger. Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, 111(16):5842–5847, 2014.

[3] Östen Dahl. Typology of sentence negation. *Linguistics*, 17(1-2):79–106, 1979.

[4] Michael Israel. Polarity sensitivity as lexical semantics. *Linguistics and philosophy*, pages 619–666, 1996.

[5] Otto Jespersen. *Negation in English and other languages*. Høst, Copenhagen, 1917.

[6] William Ladusaw. *Polarity sensitivity as inherent scope relations*. PhD thesis, University of Texas at Austin, 1979.

[7] R Core Team. R: A language and environment for statistical computing, 2018.

minimum	near-minimum	near-maximum	maximum
			

Table 1: Scenarios in Exp 1. Illustration of four scale points (degrees) for the gradable property 'transparent', which is expressed by the predicate *pleet* in the experiment.

(a) Experiment 1 (see Fig. 1)

Scenario	SP	SN	Neg-NPI
Minimum degree	✗	✓	✓
Near-minimum degree	✗	✓	✗
Near-maximum degree	✓	✗	✗
Maximum degree	✓	✗	✗

(b) Experiment 2 (see Fig. 3)

Scenario	SP	SN	Neg-Ever
Never	✗	✓	✓
Previously	✗	✓	✗
Currently	✓	✗	✗
Always	✓	✗	✗

Table 2: Scenarios that verify non-target sentences in each experiment. The ✓sign indicates that the sentence could be used in the scenario; ✗ that it could not.

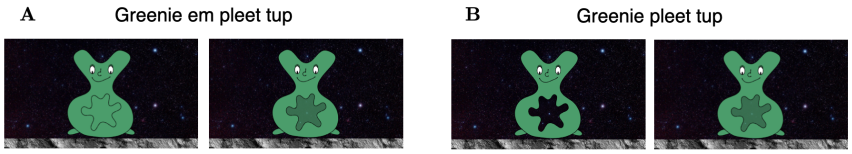


Figure 1: Sentence-picture matching trials in **(A)** Training (Neg-'At all') and **(B)** Testing (Pos-'At all') in Exp 1a.

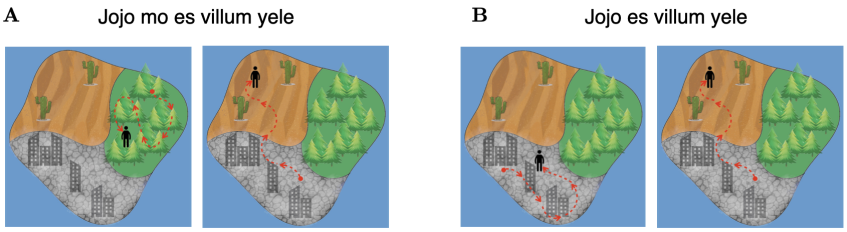


Figure 2: Sentence-picture matching trials in **(A)** Training (Neg-'Ever') and **(B)** Testing (Pos-'Ever') in Exp 2.

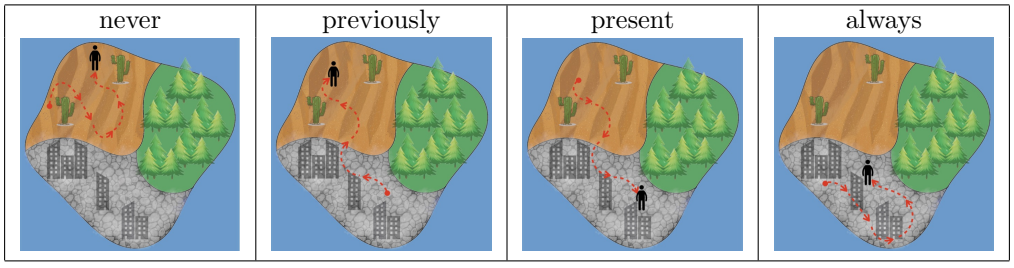


Table 3: Scenarios in Exp 2. Illustration of the four possible states for the city location, which is expressed by the predicate *villum* in the experiment.

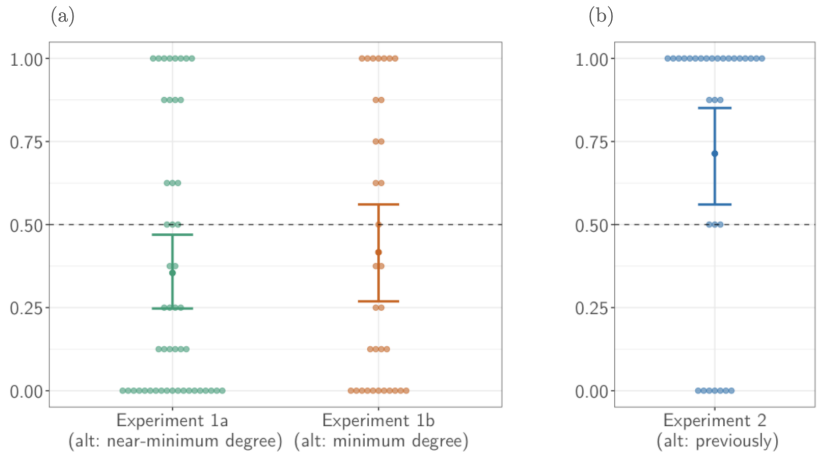


Figure 3: Results in Exps 1 and 2. (a) Proportion of 'maximum degree' responses for Pos-'At all' sentences in Exps 1a and 1b. (b) Proportion of 'always' responses for Pos-'Ever' sentences in Exp 2. Dots are individual participants, error bars are 95% bootstrapped confidence intervals.

Testing communicative and learning biases in a causal model of language evolution: A study of cues to Subject and Object*

Natalia Levshina

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
natalia.levshina@mpi.nl

1 Introduction

If we want to understand language evolution, we should understand how different cognitive and communicative biases shape the structure of human languages. This paper investigates the relationships between different linguistic cues that help to identify the grammatical roles of Subject and Object, conveying "who did what to whom". These cues include case marking, rigid word order of Subject and Object, verb-medial order, and "tight" semantics of the arguments, which is measured as association between the roles and lexemes [11]. Previous research showed correlations between case marking and word order rigidity [14][21][22], between case marking and semantic tightness [11][14], and between case marking and the position of the verb in a simple clause [8][10][11][14][22]. These correlations have been explained by language users' bias towards efficient behaviour – more exactly, saving articulation costs and avoiding additional processing costs due to reanalysis.

At the same time, there are a number of causal hypotheses about the role of sociolinguistic variables, such as population size and proportion of L2 speakers, in language evolution. One of them says that a high proportion of L2 users in a community leads to morphological simplification, such as loss of case [3][15][17]. This finding has been explained by the fact that adults are poor grammar learners [24]. In contrast, a large-scale study in [13] finds no evidence of correlations between linguistic variables and the proportion of L2 users or vehicularity as an indicator for whether a language is used by L2 speakers; instead, the study reveals a relationship between the total number of speakers and information-theoretic complexity, which represents linguistic reduncancy. The question of this paper is, how all these variables interact in a causal network and whether we find evidence of the communicative and learning biases proposed in the literature.

2 Data

In order to test these hypotheses, I used diverse sources of data:

- large web-based corpora of online news [9] annotated with Universal Dependencies;
- smaller Universal Dependencies corpora of over 100 languages [25];
- the parallel corpus of Bible translations [16] and word order data inferred from this corpus [26];

*I thank Gerhard Jäger, Matías Guzmán Naranjo and Adèle H. Ribeiro for enlightening discussions of causal models.

- the World Atlas of Language Structures (WALS) [1].

From these sources I obtained the information about three variables that help to understand "who did what to whom", according to the literature:

- entropy of Subject and Object order based on the probabilities of SO and OS orders in the corpora. The lower this measure, the more a language user can rely on word order in order to identify Subject and Object;
- whether the forms of Subject and Object are the same or distinct thanks to morphological or adpositional case marking;
- the position of the lexical verb in a transitive clause. It has been claimed that the verb-middle order helps the addressee to identify Subject and Object [8].

As for the sociolinguistic variables, I took information about the population size and L2 speaker proportions from the Ethnologue database, as well as the datasets used in [23] and [3]. The genealogical and geographic information (linguistic genus and macroarea) was also included, in order to control for the genealogical and areal dependencies between the languages. I was able to obtain the linguistic and sociolinguistic data for 112 languages from 45 genera.

3 Method

In order to infer and test causal relationships, we would need information about the past states of languages. Unfortunately, it is very often missing (but see [2] on evolution of word order in Romance languages). A popular alternative is artificial miniature language learning and communication, which has shown correlations between some of the variables of interest [6][19].

This paper develops an alternative method, namely, causal inference based on synchronic data, whose potential for linguistics has been demonstrated in [4],[20] and other works. The causal graphs were created with the help of the Fast Causal Inference (FCI) algorithm implemented in the `pcalg` R package [12] (see a linguist-friendly introduction in [5]). Since the data were mixed (numeric and categorical), I wrote additional R scipts to test conditional independence between the variables. The scripts performed likelihood ratio tests on mixed-effects regression models (more exactly, Gaussian, logistic and beta regression) to determine whether or not the variables are conditionally independent. The genealogical dependencies between the languages were controlled by treating the genera as random intercepts. The geographic macroareas were treated as fixed effects, due to the fact that there were only four areas.

4 Results

The resulting causal network is displayed in Figure 1. Causal relationships were identified only between the three of the five variables: the number of users (`Total_Users`), entropy of Subject and Object order (`SO_Entropy`) and the position of the verb (`Verb`). The cirles at the end of the arrows represent uncertainty. For example, there is a possible causal effect of `Total_Users` on `SO_Entropy`, but it can also be a bidirectional relationship (which is not likely in our case). The same holds for `Verb` and `SO_Entropy`.

Contrary to expectations, there is no support for the causal links between between the proportion of L2 users (`L2Prop`) and the linguistic variables. Surprisingly, we also find no links between the same or different forms of Subject and Object (`SO_Form`), which was strong in previous studies [14]. The Area does not play a role, either.

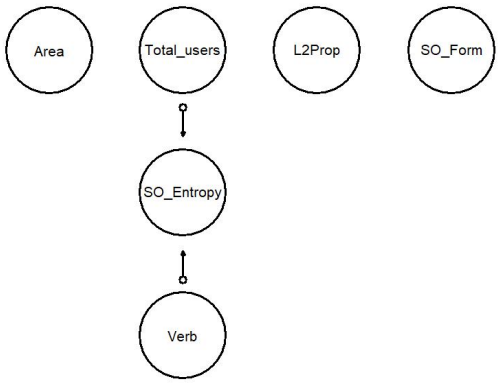


Figure 1: Causal Network

5 Discussion

The causal analysis does not support the hypothesis about the L2-related learning biases. There is no effect of proportion of L2 speakers on the main cues that help language users to distinguish between Subject and Object. Instead, the model comes to a conclusion similar to the one in [13], namely, that the total number of speakers is more likely to be relevant for language structure. In the present study, the number of speakers has an effect on how variable or fixed the order of Subject and Object is. This conclusion is intuitive. Two dominant languages, English and Chinese, have both a huge number of speakers, and very little variability in the order of Subject and Object.

How can one explain these findings? It is possible that a large number of language users increases variability of their language and cultural background and therefore increases noise in the communication channel, making the transmission of information less reliable. In this case, fixation of word order can be a response to the noise (cf. [8]). Rigid word order may be the most obvious strategy for expressing "who did what to whom" unambiguously. As for the link between the verb position and Subject - Object entropy, it is possible that rigidification helps best when the verb is in the middle.

The effect of the number of speakers on language structure has been discussed in [18] and [19]. According to the latter, larger groups of interacting participants, who learned an artificial language, developed more systematic languages over time. The increase in structure can be interpreted as a compensation for the greater linguistic variability in the larger groups. However, it is important to remember that the degree of systematicity, or language structure, was measured in that study as the correlations between string distances in the participants' linguistic output and semantic distances between the stimuli, which differed along some physical dimensions. The measure of language structure reflected morphological isomorphism and compositionality. The present study finds a causal relationship between the number of speakers and word order variability, which is a different aspect of systematicity. This is a new finding.

There are several caveats that need to be mentioned. The lack of effect of the number of speakers on case marking as one of the most salient aspects of morphological complexity may seem unexpected. But since the relevant variable in this study only describes whether the forms of Subject and Object are identical or not (including adpositional markers), perhaps this may not be the best measure for assessing the effect of L2 speakers on the overall grammatical

complexity of a language. Another problem is that the sociolinguistic variables used in this study describe the current situation, and not the previous stages, during which the grammar was shaped up. Unfortunately, historical sociolinguistic data are very difficult to find. Yet another potential factor that may affect the results is phonological complexity, which has been shown to correlate both with morphosyntactic complexity and sociolinguistic variables [7]. Nethertheless, I hope that the causal inference method presented in this study will help us to understand better how diverse cognitive and communicative biases drive language evolution.

References

[1] The world atlas of language structures online. http://wals.info, 2013.

[2] Brigitte M. Bauer. Word order. In P. Baldi and Pierluigi Cuzzolin, editors, *New Perspectives on Historical Latin Syntax. Vol 1: Syntax of the Sentence*, pages 241–316. Mouton de Gruyter, Berlin, 2009.

[3] Christian Bentz and Bodo Winter. Languages with more second language learners tend to lose nominal case. *Language Dynamics and Change*, 3:1–27, 2013.

[4] Damian E. Blasi and Sean G. Roberts. Beyond binary dependencies in language structure. In N.J. Enfield, editor, *Dependencies in Language*, page 117–128. Language Science Press, Berlin, 2017.

[5] Johannes Dellert. *Information-Theoretic Causal Inference of Lexical Flow*. Language Science Press, Berlin, 2019.

[6] Maria Fedzechkina, Elissa L. Newport, and T. Florian Jaeger. Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive Science*, 41:416–446, 2016.

[7] Gertraud Fenk-Oczlon and Jürgen Pilz. Linguistic complexity: Relationships between phoneme inventory size, syllable complexity, word and clause length, and population size. *Frontiers in Communication*, 6:626032, 2021.

[8] Edward Gibson, Steven T. Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, and Rebecca Saxe. A noisy-channel account of crosslinguistic word order variation. *Psychological Science*, 24:1079–1088, 2013.

[9] Dirk Goldhahn, Thomas Eckart, and Uwe Quasthoff. Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 759–765, Istanbul, 2012.

[10] Joseph Greenberg. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, editor, *Universals of Grammar*, pages 73–113. MIT Press, Cambridge, MA, 1966.

[11] John A. Hawkins. *A Comparative Typology of English and German. Unifying the contrasts*. Croom Helm, London, 1986.

[12] Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H. Maathuis, and Peter Bühlmann. Causal inference using graphical models with the r package pcalg. *Journal of Statistical Software*, 47:1–26, 2012.

[13] Alexander Koplenig. Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science*, 6:181274, 2019.

[14] Natalia Levshina. Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology*, 12:648200, 2021.

[15] Gary Lupyan and Rick Dale. Language structure is partly determined by social structure. *PLoS One*, 5:e8559, 2010.

[16] Thomas Mayer and Michael Cysouw. Creating a massively parallel bible corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, page

3158–3163, Reykjavik, 2014.

[17] John McWhorter. *Linguistic simplicity and complexity: Why do languages undress?* de Gruyter Mouton, Berlin, 2011.

[18] Daniel Nettle. Social scale and structural complexity in human languages. *Philosophical Transactions of the Royal Society B*, 367:1829–1836, 2012.

[19] Limor Raviv, Antje Meyer, and Shiri Lev-Ari. Larger communities create more systematic languages. *Proceedings of the Royal Society B: Biological Science*, 286:20191262, 2019.

[20] Sean G. Roberts, Anton Killin, Angarika Deb, et al. Chield: the causal hypotheses in evolutionary linguistics database. *Journal of Language Evolution*, 5:101–120, 2020.

[21] Edward Sapir. *Language: An Introduction to the Study of Speech.* Harcourt, New York, 1921.

[22] Kaius Sinnemäki. Word order in zero-marking languages. *Studies in Language*, 34:869–912, 2010.

[23] Kaius Sinnemäki and Francesca Di Garbo. Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology*, 9:1141, 2018.

[24] Peter Trudgill. *Sociolinguistic typology: Social determinants of linguistic complexity.* Oxford University Press, Oxford, 2011.

[25] Daniel Zeman et al. Universal dependencies 2.10. http://hdl.handle.net/11234/1-4758, 2022.

[26] Robert Östling. Word order typology through multilingual word alignment. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, page 205–211, Beijing, 2015.

Interpretations of plurality in personal pronouns differ across person categories: An experimental study

Mora Maldonado[1] and Carmen Saldana[2,3]

[1] Laboratoire de Linguistique de Nantes, CNRS/Nantes Université
mora.maldonado@univ-nantes.fr
[2] Department of Comparative Language Science, University of Zurich
[3] Center for the Interdisciplinary Study of Language Evolution, University of Zurich
carmen.saldanagascon@uzh.ch

1 Introduction

Personal pronouns refer to individuals defined by the communicative context, in which there are at least a speaker and an addressee, and potentially other individuals who play no active role in the conversation. The person category of a given personal pronoun specifies the main referent of the pronoun, (a.k.a. its *focal referent*). For example, singular first person pronouns (1SG) refer to the speaker in the context, singular second person pronouns (2SG) to the addressee and singular third person pronouns (3SG) to someone who is not actively participating in the conversation. When it comes to the plural forms of these pronouns, an interesting observation emerges, since plural pronouns do not seem to be interpreted in exactly the same way across person categories.

First and second plural pronouns (1PL and 2PL), such as English *we* or *you* (or *youse* or *y'all*), are not the plural of their singular alternatives like *dogs* is the plural of *dog*: Unlike these plural nouns, 1PL and 2PL pronouns do not necessarily refer to a *collection* of speakers or addressees (henceforth, *additive meaning*). Rather, they refer to a group of people that *includes* the speaker or addressee (henceforth, *associative meaning*) [4, 13, 5, 7]. For example, the pronoun *we* does not refer to multiple people speaking in unison (except in some rare cases, i.e., *chorus we*), but it is used to refer to groups whereof the speaker is part, together with other people; and the plural *you* might sometimes refer to a group containing multiple addressees (i.e., an audience) and sometimes to a single addressee plus associates. That is, both 1PL and 2PL pronouns seem to be compatible with both additive and associative meanings.

Things are different with plural third person pronouns (3PL). For a plural pronoun to be morphologically third person, it generally needs to be the case that *no* member of the group is actively participating in the conversation. Just like standard nominal plurals[1], 3PL pronouns normally refer to a homogeneous group, a collection of individuals who are neither speakers nor addressees. Thus, 3PL pronouns seem to typically have additive meanings.[2]

Altogether, there seems to be a tight relation between the interpretation of plurality in pronouns and their person category. However, there is hitherto no experimental evidence to

[1] We treat standard nominal plurality as having an additive interpretation, however, associative nominal plurality also exists for some categories of nouns in a rather large subset of the world's languages [4, 13, for an in-depth discussion]. Around half of these languages even have different markers for associative and additive meanings [8]. This is not the case for pronouns: No known language has a specific form to refer to a homogeneous group of speakers or addressees, different from a form denoting the heterogeneous, associative meaning [9, 14, 5].

[2] It has been argued that 3PL pronouns can also have associative readings, referring to a heterogeneous group of one focal non-participant plus associates [7, 18]. Teasing homogeneous and heterogeneous readings apart in this case requires a way of distinguishing between non-participants and associates, which is not entirely obvious in most uses.

confirm it. Here, we provide the first experimental investigation on speakers' interpretation of plurality in personal pronouns. We do this via an online experiment where we ask participants to choose between associative and additive meanings of plural person pronouns.

Before we move onto presenting our experiment, and in order to better frame our experimental hypotheses, we first develop further assumptions in semantic theory regarding the interpretations of plurality in personal pronouns conditioned on their person category.

1.1 The semantics of personal pronouns

Standard semantic accounts of personal pronouns [16, 11] derive these observations by assuming that 1PL and 2PL pronouns refer to a group with one focal referent plus some underspecified, additional associate(s), whereas 3PL refer to an homogeneous group of non-participants. More concretely, these proposals are built on two main assumptions. First, following common assumptions in morphological theory [14, 1], personal pronouns are defined by their reference to just two conversational roles: speaker and addressee. Second, personal pronouns only require the *inclusion* of communicative roles in the person referent: A first person pronoun (either singular or plural) requires the speaker(s) in the context to be *part of* the individual denoted, but not to be *the* individual itself. The second person imposes the same requirement but with respect to the addressee(s). A simplified denotation for each person category is given in (1), following the implementation in [12], where person and number requirements are presuppositional.

(1) a. $[\![1]\!]^c = \lambda x_e : x$ includes the speaker(s) in c. x
b. $[\![2]\!]^c = \lambda x_e : x$ includes the addressee(s) in c. x
c. $[\![3]\!]^c = \lambda x_e : x$ excludes the speaker(s) and the addressee(s) in c. x

(2) a. $[\![\text{SG}]\!]^c = \lambda x_e : |x| = 1.\ x$
b. $[\![\text{PL}]\!]^c = \lambda x_e : |x| > 1.\ x$

These assumptions, together with an intuitive semantics for number marking (as in 2), give the desirable results for first and second person pronouns. In the singular, the only way for an individual to have a cardinality of one and to include the speaker/addressee is *to be* the speaker/addressee. For 1PL and 2PL pronouns, the requirement is weaker: the group denoted by the pronoun should simply include the speaker(s) or addressee(s). Associative interpretations arise when there is a single speaker/addressee in the context, whereas additive meanings, such as the *chorus we*, arise whenever there are multiple speakers/addressees.

For third person pronouns, the only requirement is to *not* refer to speaker and addressee, given that only two roles (speaker and addressee) are assumed—but see [15]. As a result, 3PL are necessary homogeneous with respect to person roles, and end up looking more similar to plural nouns which only have additive interpretations.

By relying on the semantics of person, the account above generates both associative and additive meanings for 1PL and 2PL pronouns. However, this theory does not say why associative interpretations of 1PL and 2PL pronouns are intuitively preferred over additive interpretations, and even in many cases the only available reading (in particular for 1PL pronouns).[3]

[3]It has been alternatively proposed that plurality has a different meaning when it applies to pronouns than when it applies to nouns [6, 18]. Different approaches vary in their implementation, but the general idea is that the semantics of plural marking when applied to pronouns is inherently associative (i.e., it doesn't mean 'more than one' but rather 'plus associates'). These approaches differ substantially from the person-based proposal above. However, the asymmetry in the interpretation *across* plural personal pronouns is left partly unexplained in the same way, and we thus will not discuss these theories here.

To account for the asymmetries between the different interpretations of plurality across person categories, some authors have constrained the possible meanings of 1PL and 2PL pronouns by ruling out the existence of multiple speakers and addressees in the ontology [10]. Since there can only be one speaker/addressee in the context, the semantics in (1) can be kept as it is and additive readings will be fully blocked. A more widespread view is that preference for associative meanings is something that pertains to world knowledge: e.g., the *chorus we* reading is dispreferred because it's extremely infrequent to encounter situations with multiple speakers in the world. Interestingly, leaving interpretative preferences to world knowledge allows to account for an additional intuition about the interpretation of 1PL and 2PL pronouns, namely that the later gives rise to more additive readings than the former [19, 7, for discussion].

In order to understand how functional pressures, such as world knowledge, may interact with morphosemantic knowledge, we need a better estimation of the relation between the interpretation of plurality in pronouns and their person category. This study aims to tackle this question by providing the first quantitative investigation on speakers' interpretation of plurality in personal pronouns. We assess the hypothesis that the likelihood of interpreting a plural pronoun additively depends on the person category in a way that follows the hierarchy: 1PL $<$ 2PL $<$ 3PL. That is, the likelihood of interpreting a 1PL as meaning 'multiple speakers' is lower than the likelihood of interpreting a 2PL as meaning 'multiple addressees', which is in turn lower than the likelihood of interpreting a 3PL as 'multiple non-participants'.

2 Methods

We test the aforementioned hypothesis via two online experiments where we ask Spanish-speaking participants to choose between associative-like and additive-like meanings of Spanish personal pronouns. Experiment 1 (Exp 1) compares the interpretation of 1PL and 2PL pronouns, and Experiment 2 (Exp 2) compares the interpretation of 2PL and 3PL pronouns. Both experiments use an analogous methodology. All experimental materials, data, and analysis reported here are available at osf.io/fbnc7. The preregistered design and analysis plan is accessible at osf.io/sdtb8. The study was approved by the Ethics Committee of the School of Philosophy at the University of Zurich (Authorisation Nr. 22.6.11).

2.1 Participants

118 Spanish-speaking adult participants (59 per experiment) were recruited on Prolific. All participants were located in Spain. They were paid £1 for a five-minute experimental session.

2.2 Materials

Participants were told they would see themselves at a party (depicted as a cartoon character) chatting to acquaintances (also depicted as cartoon characters of different colours). Participants in the experiment were assigned an avatar for themselves (e.g., the white one). In each trial, participants were presented with a GIF depicting the situation at the party, and were provided with a sentence that some of the other characters uttered to them.

The GIFs displays three "focal" characters which remain still: the participant' character (which given the scenario acts as a *focal addressee*), a character standing close to it and directing its gaze towards it (*focal speaker*), and a character further away from these two, apparently not participating in the interaction (*focal other*). Each of these stationary characters was joined by a moving partner character who stood next to them for 80ms and could then be interpreted

as a non-focal, additional speaker, addressee or other. There were two variants of the GIF, differing on the distribution of the coloured characters in the scene. An example of a GIF is available here.

All sentences involved overt subject personal pronouns, for a subset of person-number combinations in Spanish. Exp 1 involved 1PL and 2PL pronouns (*nosotros* vs. *vosotros*), and Exp 2 involved 2PL and 3PL pronouns (*vosotros* vs. *ellos*). Half of the trials involving these sentences were our critical trials, the other half were used as fillers (see Section 2.3). Singular pronouns (*yo*, *tú*, *él*) were uniquely used as filler trials in both experiments. Sentences for first person pronouns were declarative statements (e.g., 'I like Rosalía'), whereas sentences for second and third person pronouns were interrogative statements (e.g., 'Do you like Rosalía?'). The predicates involved in these sentences always denote habitual activities, which in the context of a party could easily refer to characters present or absent in the scene.

2.3 Procedure

The procedure was minimally different between experiments. In Exp 1, after being presented with a GIF-sentence pair, participants had to decide: (1) who uttered the sentence (Q-SPEAKER), (2) who was the sentence addressed to (Q-ADDRESSEE), and (3) what character or group of characters the pronoun in the sentence referred to (Q-REFERENCE). All these questions involved a forced-choice amongst two alternatives, and were presented consecutively.

In Q-SPEAKER and Q-ADDRESSEE, participants had to decide whether there was a single speaker/addressee (focal referent alone) or multiple speakers/addressees by selecting amongst two descriptions. Descriptions indicated the characters by naming their color. For example, given the scenario depicted in Fig. 1, participants would have to indicate whether the sentence was uttered by the red character alone or by the red and the blue characters together.[4]

The Q-REFERENCE depended on the specific type of trial (critical or filler). In critical trials, participants had to choose between two images: a *homogeneous* picture, which displayed two characters who could arguably be assigned the same person role and a *heterogeneous* picture, which displayed one character present in the context and one character absent in the context. An illustration of these two alternatives for Q-REFERENCE is provided in Fig. 1.

The experiment involved four critical trials, two per plural pronoun. We also included 12 filler trials including singular pronouns, and four filler trials including plural pronouns but with different alternatives in Q-REFERENCE than those in the critical trials. In filler singular trials, alternatives for Q-REFERENCE involved choosing between pictures with a single character, which could be focal or not. In filler plural trials, participants were asked to choose between one heterogeneous or homogeneous picture, and a *combined* picture, which displayed all the relevant characters together (focal, non-focal and absent in the scene) together. The description of options for fillers can be found at osf.io/2bzvm. Trials for each pronoun involved one of four different predicates; the same pronoun would thus appear in four different sentences. The same four predicates were used with the two plural pronouns, and other four different predicates were used for the trials involving singular pronouns.

The combination of answers to Q-SPEAKER, Q-ADDRESSEE and Q-REFERENCE in critical trials allow us to discern between additive and associative readings. 1PL pronouns are considered to be interpreted additively if participants first consider that there are multiple speakers (in Q-SPEAKER) and then choose an homogeneous picture as the reference for the pronoun (in Q-

[4]To prevent automatic responses, where participants always select e.g., the focal referent alone, in filler trials the single speaker/addressee description could involve the non-focal referent alone (in the example, the blue character alone) rather than the focal one.

Reference). Likewise, 2pl pronouns are interpreted additively whenever participants select multiple addressees in Q-Addressee and later a homogeneous picture in Q-Reference.

Exp 2 was identical to Exp 1 except that participants were not asked the Q-Addressee. The reasoning behind this modification was that the Q-Addressee could boost additive readings of 2pl pronouns by making them available to participants (we will discuss this further in Section 3). We thus excluded this question to be able to compare directly 2pl and 3pl pronouns. Table 1 summarises the alternative responses per experiment and question type.

3 Results

Fig. 2 shows the proportion of responses compatible with an additive interpretation for each experiment and person category. For Exp 1, a response is considered to be compatible with an additive interpretation when participants first select multiple speakers or multiple addressees for 1pl (in Q-Speaker) and 2pl pronouns (in Q-Addressee) respectively, and later select homogeneous pictures (in Q-Reference).[5] In Exp 2, responses are considered to be compatible with an additive interpretation when participants select homogeneous pictures in Q-Reference.

We use *R*'s *brms* package [2] as an interface to *Stan* [3] to run Bayesian binomial mixed-effects regression models predicting the additive-like responses of plurality by person. We run two separate models, one for Exp 1 and another for Exp 2, with the same model structure. Our dependent variable is participants' responses compatible with an additive interpretation as explained above (coded as 1 if compatible, and 0 if not). We apply treatment coding to our categorical predictor of person. In Exp 1, we compare 2pl to 1pl trials; in Exp 2, we compare 3pl to 2pl. We include random intercepts for participants as well as by-participant random slopes for the effect of person. We set the same student-t prior on our fixed effect as well as on the intercept ($DF = 6, \mu = 0, \sigma = 1.5$); for the random effects, we set a half-Cauchy prior with scale parameter 10.

Results from our models suggest that, in accordance with our hypothesis, participants are more likely to choose additive-like responses for 2pl pronouns than for 1pl in Exp 1 ($\hat{\beta} = 2.574$, 90%CI $= [1.705, 3.617]$, $SE = 0.589$, $P(\hat{\beta} > 0) = 1$) and for 3pl pronouns than for 2pl pronouns in Exp 2 ($\hat{\beta} = 1.848$, 90%CI $= [0.911, 2.939]$, $SE = 0.625$, $P(\hat{\beta} > 0) = 0.99$).

In the results in Fig. 2 we nonetheless observe that the proportion of additive responses for 2pl notably varies between experiments. We believe this is due to a sort of *priming* effect, whereby in Exp1, the Q-Addressee influences subsequent Q-Reference responses in 2pl trials, boosting additive-like interpretations. Recall that this was the reason why we excluded these trials in Exp 2, as it would not have allowed us to compare 2pl and 3pl.

In order to explore this effect, we look at the proportion of multiple addressee choices in Q-Addressee questions across 1pl and 2pl in Exp 1. The idea is that participants might be more likely to conceive multiple addressees when they are specifically asked about the addressee(s) of 2pl sentences, possibly because they find it hard to distinguish between the addressee(s) and the pronominal referents. Indeed, in Fig. 3 we observe that participants choose multiple addressees with a grand mean probability of $P = 0.44$ in 1pl trials, but for 2pl trials, they select them with $P = 0.78$. After the selection of multiple addressees in 2pl trials, the only possible compatible choice in Q-Reference for critical trials is the homogeneous image: the alternative heterogeneous image does not contain both of the chosen addressees and selecting it would thus be incongruous. Altogether, this suggest that the use of 2pl pronouns in sentences is

[5] We exclude trials in Exp 1 where participants select multiple speakers or addressees in Q-Speaker or Q-Addressee and later do not select pictures that include these referents in Q-Reference. These are considered to be incongruous (errors). We excluded the responses of 5 (out of 118) 1pl and 7 (out of 118) 2pl trials.

indeed priming participants to select multiple addresses more than they would do for sentences containing another pronoun (i.e., 1PL), and it is inevitably boosting the choice of homogenous images in Q-REFERENCE, which it is not the case in Exp 2.

In our confirmatory results (Fig. 2) we also observe that, despite its rarity, participants are equally likely to select multiple speakers in 1PL trials than single speakers (which is the norm). This can be also due to the difficulty of teasing apart speakers from pronominal referents in 1PL trials, the same way we observe it is hard to distinguish addressees and referents in 2PL trials. Fig. 4A shows that participants choose multiple speakers in Q-SPEAKER notably more often for 1PL than for 2PL trials, where they select singular speakers almost exclusively (as in 2PL and 3PL trials in Exp 2 as well, see Fig. 4B).

4 Discussion

Our study provides the first experimental evidence confirming that the likelihood of additive interpretations of plural pronouns depends on the person category. Additive readings are more likely in 3PL than in 2PL pronouns (Exp 2), and in 2PL pronouns than in 1PL pronouns (Exp 1). This is consistent with our hypothesised hierarchy: $1 < 2 < 3$.

Although intuitive, the difference in interpretation between 1PL and 2PL pronouns is not accounted for by most theories of person—but see [19]. Our results provide evidence in favour of an account whereby the source of this asymmetry is not due to a difference in the semantics of these pronouns alone, but rather to the combined influence of world knowledge and semantics: the existence of multiple speakers is less likely than the existence of multiple addressees (independently of language). Exp 1 moreover suggests that additive interpretations are *possible* for both 1PL and 2PL pronouns, indicating that the person ontology should allow for multiple speakers and addressees—*contra* [10]. Interestingly, research on the cross-linguistic distribution of person systems has argued that some morphological patterns (e.g., the likelihood of person-number suppletion) also seem to obey a $1 < 2 < 3$ hierarchy [17], suggesting that the interaction of morphosemantic knowledge and functional pressures not only influences the preference of a certain interpretation but also finds its way into the grammar.

Note, however, that even though the results from Exp 2 confirm our hypothesis, we cannot establish an additive interpretation for what we have been calling 'additive-like' responses in Q-REFERENCE. Unlike for Exp 1, here we do not know whether participants conceive multiple or single addressees when they select homogeneous images, and thus we are inflating additive interpretations by only taking into account Q-REFERENCE responses. However, even with this conservative approach, we still observe the predicted $2 < 3$ asymmetry.

Surprisingly, we also found comparable results for 2PL in Exp 1 and 3PL in Exp 2 (see Fig. 2). This could suggest a hierarchy $1 < 2/3$ (rather than $1 < 2 < 3$) with regards to plural interpretation. However, we also observe a notable difference between additive responses for 2PL across experiments, which we have argued results form the priming effect the Q-ADDRESSEE has on the selection of multiple addressee responses in communicative contexts where the sentence uttered contains a 2PL pronoun. We showed that participants are indeed more likely to select homogeneous images in Q-REFERENCE in Exp 1 than in Exp 2, indicating that a direct comparison between Exp 1 and 2 is hard to make.

References

[1] Jonathan David Bobaljik. Missing persons: A case study in morphological universals. *The Linguistic Review*, 25(1–2):203–230, 2008.

[2] Paul-Christian Bürkner. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, 10(1):395–411, 2018.

[3] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[4] Greville G Corbett. *Number*. Cambridge University Press, 2000.

[5] Michael Cysouw. *The Paradigmatic Structure of Person Marking*. OUP Oxford, 2003.

[6] Michael Cysouw. The expression of person and number: A typologist's perspective. *Morphology*, 21(2):419–443, 2011.

[7] Michael Daniel. Associative plural as indexical category. *Language Sciences*, 81:101256, 2020.

[8] Michael Daniel and Edith Moravcsik. The associative plural. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013.

[9] Joseph H Greenberg. The first person inclusive dual as an ambiguous category. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 12(1):1–18, 1988.

[10] Daniel Harbour. *Impossible Persons*. Linguistic Inquiry Monographs. MIT Press, 2016.

[11] Irene Heim. Features on bound pronouns. In Daniel Harbour, David Adger, and Susana Béjar, editors, *Phi theory: phi-features across modules and interfaces*, chapter 2. Oxford University Press, Oxford, 2008.

[12] Angelika Kratzer and Irene Heim. *Semantics in generative grammar*, volume 1185. Blackwell Oxford, 1998.

[13] Edith A Moravcsik. A semantic analysis of associative plurals. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 27(3):469–503, 2003.

[14] Robert Rolf Noyer. *Features, positions and affixes in autonomous morphological structure*. PhD thesis, Massachusetts Institute of Technology, 1992.

[15] Uli Sauerland. On the semantic markedness of phi-features. In Daniel Harbour, David Adger, and Susana Béjar, editors, *Phi theory: phi-features across modules and interfaces*, chapter 3. Oxford University Press, Oxford, 2008.

[16] Philippe Schlenker. Person and binding (a partial survey). *Italian Journal of Linguistics/Rivista di Linguistica*, 16(1):155–218, 2004.

[17] Anna Siewierska and Dik Bakker. Suppletion in person forms: The role of iconicity and frequency. *Languages across boundaries: Studies in memory of anne siewierska*, pages 347–396, 2013.

[18] Jolijn Sonnaert. *The atoms of person in pronominal paradigms*. Netherlands Graduate School of Linguistics, 2018.

[19] Stephen Wechsler. What 'you' and 'I' mean to each other: Person indexicals, self-ascription, and theory of mind. *Language*, 86(2):332–365, 2010.

	Exp 1	Exp 2
Q-Speaker	(a) focal speaker (*single*) (b) focal + non-focal speakers (*multiple*)	(a) focal speaker (*single*) (b) focal + non-focal speakers (*multiple*)
Q-Addressee	(a) focal addressee (*single*) (b) focal + non-focal addressees (*multiple*)	
Q-Reference	(a) focal + non-focal (*homogeneous*) (b) focal + associates (*heterogeneous*)	(a) focal + non-focal (*homogeneous*) (b) focal + associates (*heterogeneous*)

Table 1: Alternative answers per question and experiment. Labels are indicated in italics.

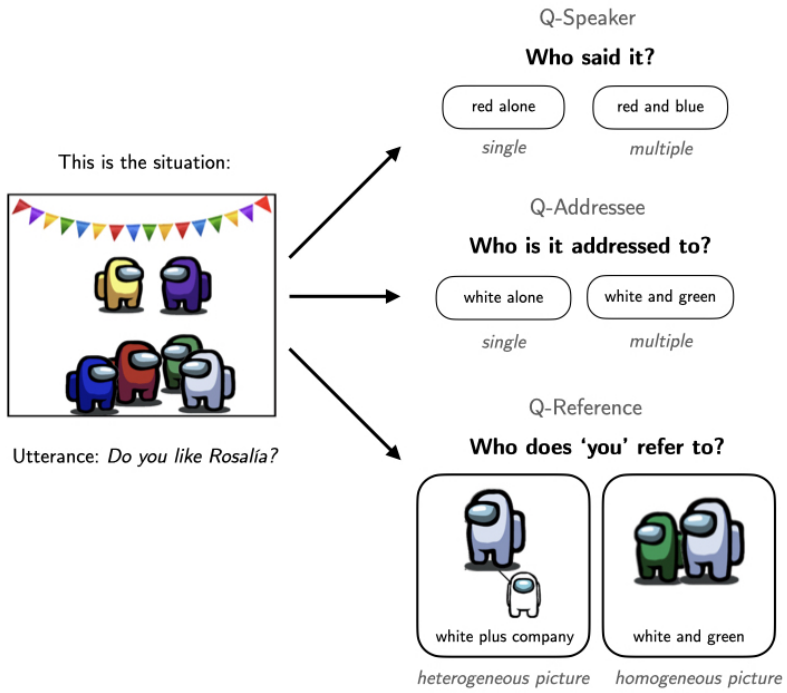


Figure 1: Schematic trial (Exp 1: 2PL pronoun). Each question was presented with the GIF-sentence pair. Original materials in Spanish can be found in osf.io/fbnc7. Labels in italics were not part of the trial.

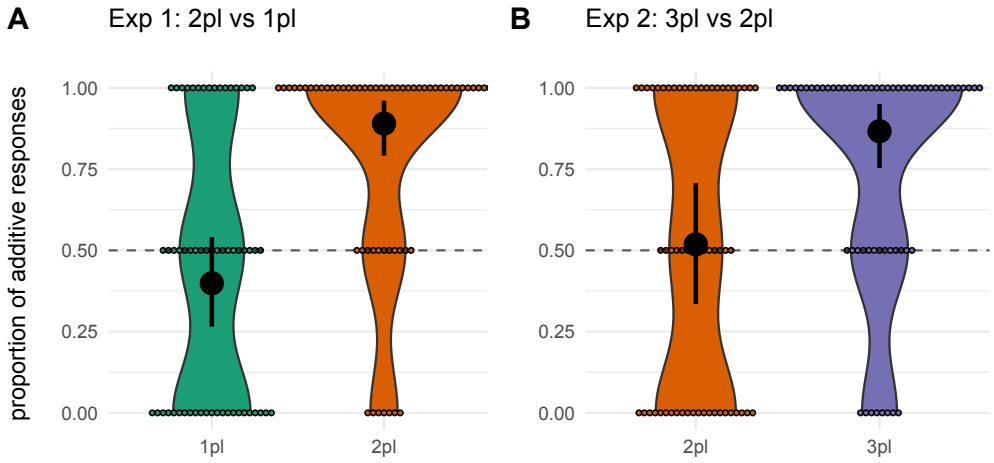


Figure 2: Proportion of additive-like responses in Exp 1 (A) and Exp 2 (B). We show the model's estimate means (black dot) and 90%CI, as well as kernel density plots indicating the peaks in the data distribution. The small coloured dots illustrate participants' individual scores. The dash line indicates the chance level.

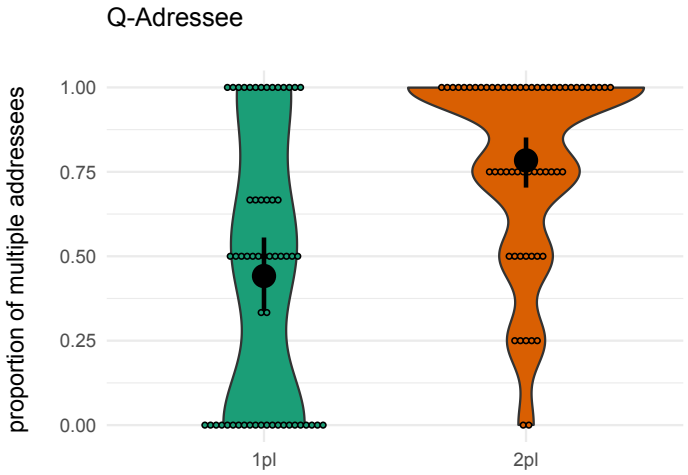


Figure 3: Proportion of selection of multiple addresses in Q-ADDRESSEE for 1PL and 2PL trials in Exp 1. We show the means and standard errors, as well as kernel density plots indicating the peaks in the data distribution. The small coloured dots illustrate participants' individual scores. Note that we include the responses of all Q-ADDRESSEE plural trials, regardless of whether they are followed by a critical Q-REFERENCE trial or not.

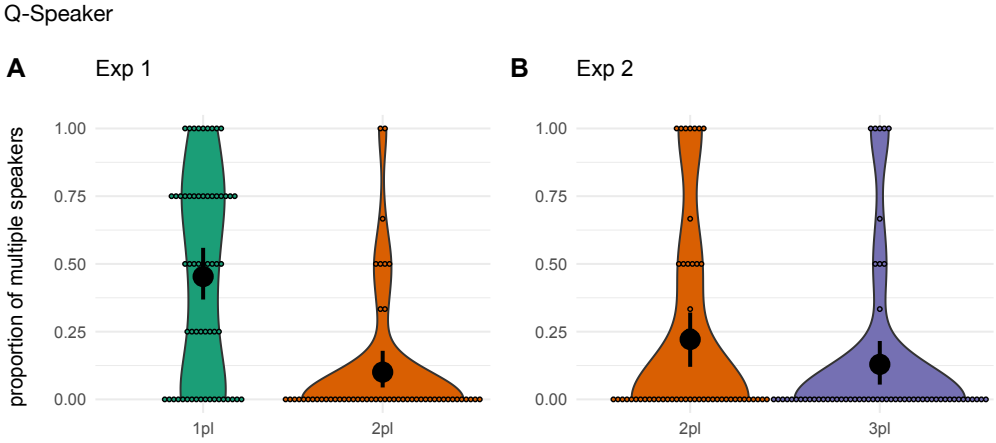


Figure 4: Proportion of selection of multiple speakers in Q-SPEAKER for 1PL and 2PL trials in Exp 1 (A), and for 2PL and 3PL trials in Exp 2 (B). We show the means and standard errors, as well as kernel density plots indicating the peaks in the data distribution. The small coloured dots illustrate participants' individual scores. Note that here we also include the responses of all Q-SPEAKER plural trials, regardless of whether they are followed by a critical Q-REFERENCE trial or not.

Cataphoric presupposition verification and symmetric filtering: The case of Occasion verbs

Oliver Bott and Torgrim Solstad

Bielefeld University
{oliver.bott,torgrim.solstad}@uni-bielefeld.de

Abstract

Presenting evidence from four offline rating experiments, we argue that a verb class we characterize as *Occasion verbs* (e.g., *thank*, *criticize*, *congratulate*; partly overlapping with *judgment verbs* [13]) displays the intriguing property of allowing for cataphoric presupposition verification more broadly than other triggers discussed in previous research. Experiment 1 through 3 used methods established by Tonhauser and colleagues ([33, 34]) to show that Occasion verbs do indeed pattern with a selection of thirteen other well-known triggers (e.g., factive and aspectual verbs, demonstrative noun phrases). In addition, Experiments 1 and 2 also provided evidence that Occasion verbs – as opposed to those well-established triggers – allow for the cataphoric verification of presuppositions in a separate clause. Experiment 4 provided more targeted evidence as to the compositional consequences of this cataphoric verifiability. We compared the filtering behaviour of Occasion verbs with factive or aspectual triggers (e.g., *know* and *continue*) for conjunctions in the antecedent of conditionals. The results show that while factive and aspectual verbs show left-to-right filtering asymmetry (cf. [22]), Occasion verbs display symmetric filtering.

1 Introduction

The expressions known as "presupposition triggers" (cf. e.g., [21]) constitute a heterogeneous class, characterized by Karttunen as the "zoo of presupposition triggers" ([18]). Thus, triggers of presuppositions have been shown to display projective variability [18, 1, 15, 12], as, for instance, reflected in the distinction between weak and strong triggers. In the present paper, we argue that triggers may also vary with regard to how easily they allow for discourse-cataphoric verification, that is, in subsequent discourse. More specifically, we will offer experimental evidence that a verb class that we call *Occasion verbs* (e.g., *thank*, *congratulate*, *praise*, *criticize*; partly overlapping with Fillmore's *judgment verbs* [13]) is more flexible than more well-established triggers like factive or aspectual verbs with regard to such cataphoric verification, thereby possibly requiring a cage of its own in the zoo of triggers. This conclusion is based on evidence from four rating studies in German, in which we, on the one hand, showed that Occasion verbs do indeed project like other triggers in the Family-of-Sentence contexts [10]. Furthermore, the results support the conclusion that the presupposed content of Occasion verbs is special in the way it relates to at-issue content, allowing both left-to-right as well as right-to-left filtering.

Presuppositions display a number of anaphoric properties, most famously formulated by van der Sandt [35], according to whom pronominal anaphora are but a special case of presupposition-triggering expressions [14]. In accordance with this idea, presuppositions have mostly been taken to be characterized by *anaphoric asymmetry* or *left-right asymmetry* (see discussion in e.g., [24, 9, 22, 3]): Just as cataphoric pronominal reference is only possible for specific constructions, presupposed material must also generally precede the presupposition trigger (unless it can be accommodated). Possible exceptions to this principle have been discussed under the notion of *postsuppositions* ([8]). For instance, in Romanian *Şi A şi B au fugit* 'A as well as B ran away'

the presupposition triggered by *şi* 'too' ("someone else than A ran away") can be verified by *B*. What is more, [9] also observed that for French *aussi* 'too', the presupposed content may follow the trigger. However, the availability of cataphoric presupposition verification has been taken to be strongly limited. Whereas postsuppositions are limited to sentence-internal verification, *aussi* requires the two sentences to be conjoined via *or*, *if* or *unless* (see also [7]).

Experimental research has provided mixed evidence as to the possibility of symmetrical filtering. Beyond the study by [9], [6] found the cataphoric verification of presuppositions to be marginally better for *if* conditionals as opposed to *and* conjunctions for the trigger *again*. In yet another experimental study, however, Mandelkern et al. [22] investigated filtering for conjunctions in the antecedent of conditionals, arguably a more constrained configuration for investigating presupposition projection (*If Mary used to do Jivamukti yoga and she stopped doing yoga, then Matthew will interview her for his story*). Employing a discourse rating paradigm in their Experiment 3, [22] found that for conjunction, only asymmetric left-to-right filtering is available, as expected on most theoretical approaches. Interestingly, though, applying the same methods as in [22], [17] found filtering to be symmetric for disjunction, that is, they found no significant difference between left-to-right and right-to-left ordering of trigger and presupposed content. In conclusion, thus, the experimental evidence – in particular the conjunction vs. disjunction studies by [22] and [17] – suggests a mixed picture. Although the results cannot be seen as having resolved the (a)symmetry, they seem to confirm what theoretical studies have proposed: If right-to-left filtering is available, it is heavily constrained.

In this paper, we present empirical evidence that *Occasion verbs* not only allow for the cataphoric verification of presuppositions, but also that they do so at a more general discourse level, without the constraints discussed above. Furthermore, we also show that they allow for right-to-left filtering in constructions parallel to those tested by [22]. While they have received little attention in (theoretical) linguistics, Occasion verbs figure prominently in psycholinguistic research on discourse expectations as the so-called Agent-Evocator subclass of Implicit Causality verbs ([2]). Occasion verbs subsume some of the well-known 'verbs of judging' [13, 23] such as *blame*, *criticise*, and *praise*, but also 'non-judgment verbs' like *thank*, or *congratulate*. On our analysis, these verbs presuppose that there is an occasion upon which the agent acts. If one thanks or criticises someone, there must be some (preceding) eventuality the agent considers worthy of gratitude or critique ([28, 29]). The cataphoric verifiability of this presupposition is evidenced by sequences such as *John thanked Lisa. [...] She had helped him on the day before*. Importantly, the cataphoric potential seems related to a discourse expectation that the eventuality denoted by the verb must be specified in subsequent discourse ([19, 4, 20]).

The present study thus provides additional evidence for the complex properties of projective content. The four experiments we report on compared Occasion verbs with other presupposition triggers with respect to their potential for cataphoric verification, their projectivity across varying contexts and their filtering behaviour.

2 Experiments 1 and 2

Experiments 1 and 2 used identical methods to test for the status of sixteen Occasion verbs as presupposition triggers as well as their discourse-cataphoricity in two different blocks. While experiment 1 ($N = 71$) compared Occasion verbs to seven well-established presupposition triggers like personal pronouns, demonstratives, possessives, factive and aspectual verbs, Experiment 2 ($N = 60$) compared Occasion verbs to six other triggers such as clefts, definite descriptions, *too* and *again*. In this proceedings paper, we will greatly gloss over particulars of the experiments. All materials and statistical analyses can be found here: `https://osf.io/9y3nu/` (see [29]).

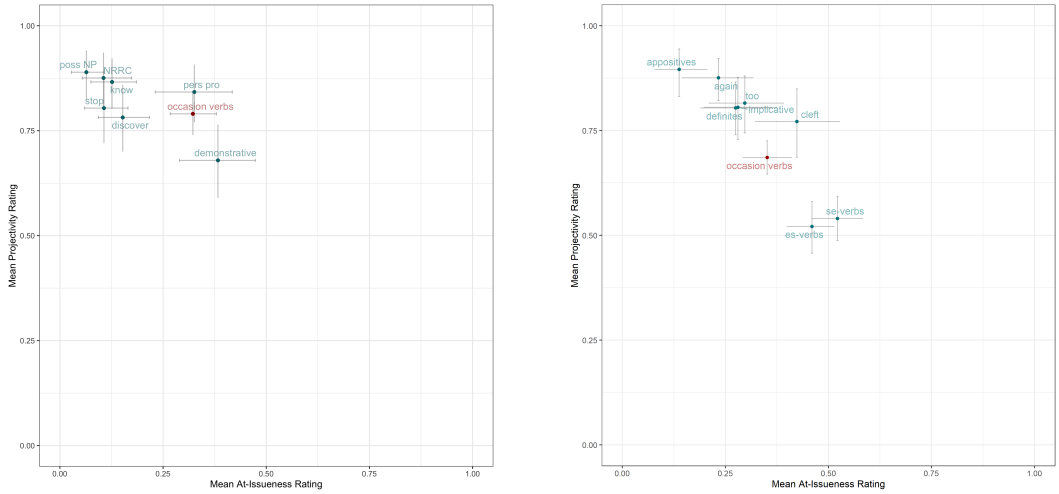


Figure 1: Mean projectivity and at-issueness ratings (and 95% CIs) for Occasion verbs as compared to other presupposition triggers in Experiments 1 (left) and 2 (right)

Projectivity and (non-)at-issueness: We will first turn to the second block of the experiments, which established that Occasion verbs do indeed pattern with other presupposition triggers not only in terms of projectivity, but also non-at-issue content, following the assumptions behind the Gradient Projection Principle [34], which states that content projects to the extent that it is non-at-issue (see [26]). We adopted the methods in [34] for eliciting projectivity and (non-)at-issueness ratings. Participants were asked to imagine overhearing utterances as part of brief dialogue snippets. Those utterances embedded presupposition triggers in questions as in: *Leo asks: "Did Ann congratulate Pete?"*. Projectivity was elicited by asking participants yes-no questions like *Is Leo certain that there was an occasion for Ann to congratulate Pete?*. If the (presumed) presupposition of *congratulate* projects, we would expect a positive answer. (Non-)at-issueness was measured via *whether*-questions such as *Did Leo ask whether there was an occasion for Ann to congratulate Pete?*. If the presupposed "occasion content" is not-at-issue, this question should evoke a negative answer. Answers were collected with a slider on a scale from 0 ("no") to 100 ("yes"). Mean projectivity and at-issueness ratings for the two experiments are presented in Figure 1.

The inference of occasioning events triggered by Occasion verbs clearly projected with an on average projectivity rating of 0.74 and a mean at-issueness rating of 0.33 over the two experiments. A clustering analysis corroborated that Occasion verbs pattern with the other presupposition triggers. As a side remark, this analysis also confirmed that non-presuppositional Implicit Causality verbs (cf. [5, 30]) do not pattern with those other triggers (called "se/es-verbs" in the right panel in Figure 1). The data importantly also replicates the finding by [34] that at-issueness and projectivity are negatively correlated (Pearson's $r = -0.68$, $p < .01$). Thus, we feel confident in concluding that Occasion verbs can be viewed as presupposition triggers, or, more generally, as "expressions associated with projective content" ([33]).

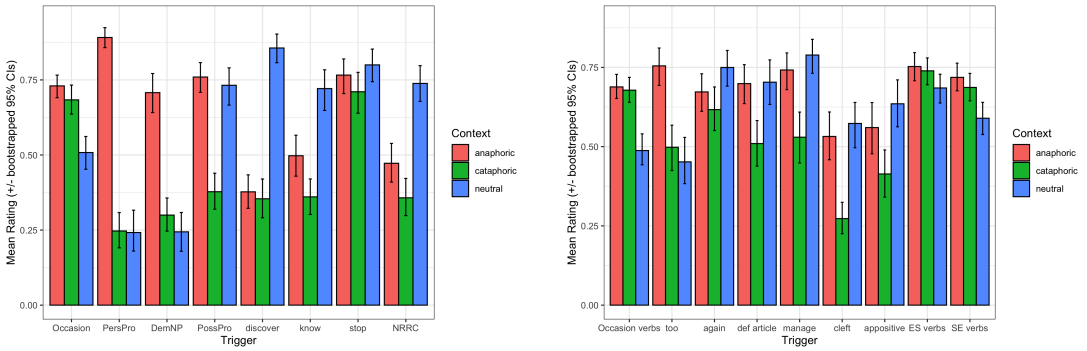


Figure 2: The discourse-anaphoric properties of Occasion verbs vs. other triggers tested in Experiments 1 (left) and 2 (right) comparing anaphoric and cataphoric to neutral conditions

Cataphoricity and anaphoricity: The other block of the two experiments investigated the availability of cataphoric (and anaphoric) verification of the presupposed content of Occasion verbs and the other presupposition triggers. The triggering expressions were embedded in brief discourses realizing three conditions:

(1) *Preceding context:* Ann and Pete are neighbours ...

a. ***m*-neutral condition:** She congratulated him.
b. **anaphoric condition:** It was his birthday yesterday. She congratulated him.
c. **cataphoric condition:** She congratulated him. It was his birthday yesterday.

The *m(eaning)-neutral condition* in (1a) (cf. [33]) required accommodation of the presupposition and served as a baseline. The *anaphoric condition* additionally included a sentence verifying the presupposition that preceded the target sentence containing the trigger, as in (1b). Crucially, in the *cataphoric condition* (1c) this additional presupposition-verifying sentence followed the target sentence. Participants provided discourse coherence ratings using a slider (0 = "completely incoherent" to 100 = "perfectly coherent"). The results are shown in Figure 2.

For reasons of space, we only compare Occasion verbs with personal pronouns and demonstrative noun phrases: In line with previous research, personal pronouns and demonstrative noun phrases required the explicit introduction of an antecedent (neutral condition; right-most, blue bar in the bar triplets). Moreover, cataphoric reference to a subsequent discourse entity was hardly possible (cataphoric condition; middle, red bar). For these two types of expressions, neutral and cataphoric conditions were rated equally bad with no statistically significant difference (pronouns: $\chi^2(1) = .03$, $p = .87$; demonstrative noun phrases: $\chi^2(1) = 1.92$, $p = .17$). By contrast, Occasion verbs allowed for both anaphoric (left-most, red bar) and – as the only trigger in the study not easily allowing for accommodation in the first place – cataphoric verification of the presupposition. For Occasion verbs, both of these conditions were rated significantly better than the neutral condition (anaphoric condition: $\beta = 22.3$, $se = 2.92$, $\chi^2(1) = 21.76$, $p < .001$; cataphoric condition: $\beta = 17.85$, $se = 2.76$, $\chi^2(1) = 36.72$, $p < .001$).

Summing up the overall results of Experiments 1 and 2, we showed that Occasion verbs are in fact presuppositional in nature. This new trigger class patterned with other, well-understood triggers with respect to projectivity and at-issueness. At the same time, Occasion verbs proved to be special in allowing for cataphoric presupposition justification.

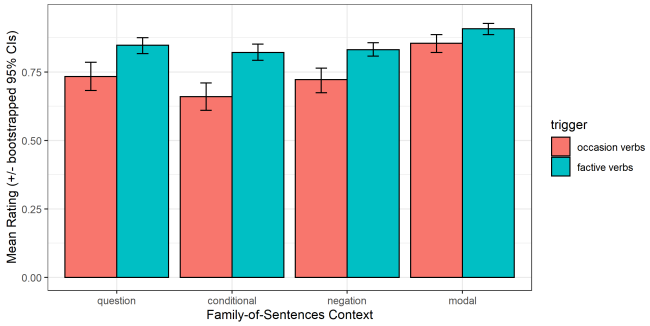


Figure 3: Projectivity of Occasion and factive verbs in Family-of-Sentence contexts (Exp. 3)

3 Experiment 3

Empirical work on factive and other kinds of sentence-embedding predicates have provided evidence for heterogeneity among triggers also with respect to Family-of-Sentence environments [27, 11, 25]. With a design similar to that used in the blocks investigating projectivity and at-issueness in the previous two experiments, Experiment 3 ($N = 105$) investigated whether the effects found in those experiments generalize to other entailment-canceling environments, more precisely to all so-called Family-of-Sentences environments [10], that is, for triggers embedded in questions, under negation, conditionals and modals. The factive verbs (*know*, *reveal*, *regret*) were included as controls.

The aggregated projectivity ratings in the four Family-of-Sentence contexts are summarized in Figure 3. The results can be summarized as follows: Even though there are slight differences between Family-of-Sentence contexts, the projective nature of the inference triggered by Occasion verbs generalizes to all Family-of-Sentence contexts. As in the previous experiments, these findings were confirmed by linear mixed-effect regression analyses.

4 Experiment 4

Experiment 4 put the cataphoric verifiability of occasion presuppositions to a stricter compositional test: Allowing cataphoric verification, occasion presuppositions should also display filtering from right-to-left in addition to the commonly assumed left-to-right filtering. As mentioned above, [22] has suggested that such filtering is strictly asymmetric under conjunction. In this experiment ($N = 58$), we therefore contrasted Occasion verbs – never tested before with respect to filtering – with the aspectual and factive presupposition triggers from [22] in a conceptual replication of their study. We used a 2 (TRIGGER TYPE) × 3 (ANTECEDENT TYPE) within design as follows: The two types of triggers appeared without or *in conjunction with* content verifying the presupposition in the antecedent of a conditional as follows:

(2) a. **Conjoined, trigger first:** If Pete congratulated Ann and she won a race, then ...
b. **Conjoined, trigger last:** If Ann won a race and Pete congratulated her, then ...
c. **trigger-only:** If Pete congratulated Ann, then ...

If filtering is symmetric, conjunct order – trigger-first (2a) vs. trigger-last (2b) – shouldn't make a difference. Under asymmetric filtering, however, trigger-first (2a) should display projection, whereas trigger-last (2b) should not, as the presupposition is filtered in this case. In the trigger-

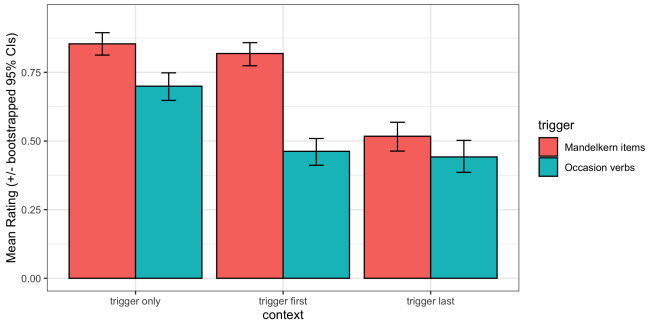


Figure 4: Left-to-right and right-to-left filtering for Occasion verbs and factive verbs in Exp. 4

only condition (2c), the presupposition should always project. The results for the projectivity ratings (again using the methods in [34]) are shown in Figure 4.

We found a clear difference between Occasion verbs and the verbs tested by [22]. More precisely, as confirmed in a linear mixed effects regression analysis, the results displayed a clear interaction with respect to the right-to-left filtering condition: For factive and aspectual verbs, the TRIGGER-FIRST and TRIGGER-ONLY conditions displayed presupposition projection, whereas the TRIGGER-SECOND condition did not, as we would expect under left-to-right filtering.

In contrast, Occasion verbs led to the orthogonal pattern, with TRIGGER-FIRST conditions patterning with TRIGGER-LAST conditions, both showing equally low projection scores (i.e., filtering). Put differently, Occasion verbs displayed perfect symmetry with respect to filtering. Our current take is that this is related to the lack of informative presuppositions (e.g., [31]) for Occasion verbs (unlike factive verbs, they are not clause-embedding).

5 Conclusion

This study shows two things: The "occasioning eventualities" associated with Occasion verbs are in fact presuppositions, sharing defining characteristics with a wide range of other triggers (Experiments 1 through 3). At the same time they prove to be special, as they were the only triggers that allowed for discourse-cataphoric verification (Experiments 1 and 2). Adding more weight to this conclusion, our study shows that these contents easily allow for right-to-left filtering, with no preference for incremental presupposition processing (Experiment 4). Interestingly, the cataphoric verifiability of this trigger type extends well beyond a single sentence boundary.

Our results speak against approaches principally upholding asymmetry, such as [16, 17], who have argued that there is a fundamental difference in the availability of filtering for conjunctions and disjunctions, with only the latter allowing it symmetrically. Rather, our results can be more easily accommodated if one assumes filtering to be symmetric, and apparent counterexamples to be infelicitous due to informational redundancy [32]. Our current take is that Schlenker ([24]) comes closest to representing such a view in characterizing left-to-right filtering as a bias arising from a preference for incremental left-to-right processing.

Summing up, the presuppositional nature of Occasion verbs, in particular, the availability of cataphoric presupposition verification, is a discourse phenomenon with interesting consequences for compositional approaches to projection. One possible upshot for the composition of presupposed content with at-issue content is that we may expect to see variation across trigger types with regard to the availability of symmetric filtering.

References

[1] Dorit Abusch. Presupposition Triggering from Alternatives. *Journal of Semantics*, 27(1):37–80, 2009.

[2] Terry Kit-fong Au. A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language*, 25(1):104–122, 1986.

[3] David I. Beaver, Bart Geurts, and Kristie Denlinger. Presupposition. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition, 2021.

[4] Oliver Bott and Torgrim Solstad. From verbs to discourse – a novel account of implicit causality. In Barbara Hemforth, Barbara Mertins, and Cathrine Fabricius-Hansen, editors, *Psycholinguistic Approaches to Meaning and Understanding Across Languages*, pages 213–251. Springer, New York, 2014.

[5] Oliver Bott and Torgrim Solstad. Discourse expectations: explaining the implicit causality biases of verbs. *Linguistics*, 59(2):361–416, 2021.

[6] Adrian Brasoveanu and Jakub Dotlačil. Incremental and predictive interpretation: Experimental evidence and possible accounts. In Sarah D'Antonio, Mary Moroney, and Carol-Rose Little, editors, *Semantics and Linguistic Theory (SALT)*, volume 25, pages 57–81. LSA and CLC Publications, 2015.

[7] Adrian Brasoveanu and Jakub Dotlačil. *Computational Cognitive Modeling and Linguistic Theory*. Springer International Publishing, 2020.

[8] Adrian Brasoveanu and Anna Szabolcsi. Presuppositional *Too*, postsuppositional *Too*. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *The dynamic, inquisitive, and visionary life of* Φ*,* $?\Phi$*, and* $\diamond\Phi$*: a festschrift for Jeroen Groenendijk, Martin Stokhof, and Frank Veltman*, pages 55–64. ILLC Publications, 2013.

[9] Emmanuel Chemla and Philippe Schlenker. Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics*, 20(2):177–226, 2012.

[10] Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar*. MIT Press, Cambridge, Mass., 1990.

[11] Marie-Catherine de Marneffe, Mandy Simons, and Judith Tonhauser. The CommitmentBank: Investigating projection in naturally occurring discourse. In M. Teresa Espinal, Elena Castroviejo, Manuel Leonetti, Louise McNally, and Cristina Real-Puigdollers, editors, *Proceedings of Sinn und Bedeutung 23*, volume 2, pages 107–124, Bellaterra, 2019. Universitat Autònoma de Barcelona.

[12] Judith Degen and Judith Tonhauser. Are there factive predicates? an empirical investigation. *Language*, 98(3), 2022.

[13] Charles J. Fillmore. Verbs of judging: An exercise in semantic description. *Paper in Linguistics*, 1(1):91–117, 1969.

[14] Bart Geurts. *Presuppositions and Pronouns*. Elsevier, Amsterdam, 1999.

[15] Jacques Jayez, Valeria Mongelli, Anne Reboul, and Jean-Baptiste van der Henst. Weak and strong triggers. In Florian Schwarz, editor, *Experimental Perspectives on Presuppositions*, pages 195–213. Springer International Publishing, Cham, 2015.

[16] Alexandros Kalomoiros. Limited symmetry. In *ESSLLI Student Session Proceedings?*, 2021.

[17] Alexandros Kalomoiros and Florian Schwarz. Presupposition projection from disjunction is symmetric. *Proceedings of the Linguistic Society of America*, 6(1):556–571, 2021.

[18] Lauri Karttunen. Some observations on factivity. *Paper in Linguistics*, 4(1):55–69, 1971.

[19] Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L. Elman. Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44, 2008.

[20] Andrew Kehler and Hannah Rohde. Prominence and coherence in a Bayesian theory of pronoun interpretation. *Journal of Pragmatics*, 154:63–78, 2019.

[21] Stephen C. Levinson. *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge University Press, 1983.

[22] Matthew Mandelkern, Jérémy Zehr, Jacopo Romoli, and Florian Schwarz. We've discovered that projection across conjunction is asymmetric (and it is!). *Linguistics and Philosophy*, 43(5):473–514, 2020.

[23] James D. McCawley. Verbs of bitching. In D. J. Hockney, W. L. Harper, and B. Freed, editors, *Contemporary research in philosophical logic and linguistic semantics*, pages 313–332. Reidel, Dordrecht, 1975.

[24] Philippe Schlenker. Local contexts. *Semantics and Pragmatics*, 2:1–78, 2009.

[25] Judith Sieker. Projective variability of factive and semifactive verbs? an experimental investigation on the influence of the family of sentences contexts on projectivity. Master's thesis, Bielefeld University, Bielefeld, 2022.

[26] Mandy Simons, Judith Tonhauser, David Beaver, and Craige Roberts. What projects and why. In Nan Li and David Lutz, editors, *Proceedings of SALT 20*, pages 309–327, 2010.

[27] E. Allyn Smith and Kathleen Curry Hall. The relationship between projection and embedding environment. In *Proceedings of the 48th Meeting of the Chicago Linguistics Society*, 2014.

[28] Torgrim Solstad and Oliver Bott. Towards a formal theory of explanatory biases in discourse. In Maria Aloni, Michael Franke, and Floris Roelofsen, editors, *Proceedings of the Amsterdam Colloquium 2013*, pages 203–210, Amsterdam, 2013. ILLC, University of Amsterdam.

[29] Torgrim Solstad and Oliver Bott. Cataphoric presuppositions: The case of occasion verbs. Unpublished manuscript., 2022.

[30] Torgrim Solstad and Oliver Bott. On the Nature of Implicit Causality and Consequentiality: The Case of Psychological Verbs. *Language, Cognition and Neuroscience*, 37(10):1311–1340, 2022.

[31] Jennifer Spenader. Factive presuppositions, accommodation and information structure. *Journal of Logic, Language and Information*, 12(3):351–368, 2003.

[32] Robert Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics*, volume 9, pages 315–322. Academic Press, New York, 1978.

[33] Judith Tonhauser, David I. Beaver, Craige Roberts, and Mandy Simons. Toward a taxonomy of projective content. *Language*, 89:66–109, 2013.

[34] Judith Tonhauser, Judith Degen, and David I. Beaver. How projective is projective content? gradience in projectivity and at-issueness. *Journal of Semantics*, 35:495–542, 2018.

[35] Rob A. van der Sandt. Presupposition projection as anaphora resolution. *Journal of Semantics*, 9:333–377, 1992.

A constraint on presupposition accommodation *

Omri Doron and Jad Wehbe

Massachusetts Institute of Technology
omrid@mit.edu jadwehbe@mit.edu

Abstract

We discuss a constraint on global accommodation proposed by Heim stating that an accommodated presupposition must not settle the QUD [Hei15]. We argue that this constraint follows from the asymmetry between the pragmatic status of presuppositions and assertions assumed by the satisfaction theory [Sta70, Kar74]. We provide evidence that local accommodation is sensitive to this constraint, thus arguing that local accommodation maintains the asymmetry. This poses a challenge to theories of local accommodation.

1 Introduction

An important question in the literature on presuppositions is how to identify their pragmatic status. Under the satisfaction view, presuppositions impose conditions on the common ground, such that a sentenced can only be uttered if its presupposition is entailed by the common ground [Sta70, Kar74]. At first glance, this seems to incorrectly predict that sentences with presuppositions can never be felicitously uttered in a context where the presupposition is not met. Consider (1) for example. Here, the presupposition that B has a car is not met in the common ground, but nevertheless B's response is felicitous.

(1) **Context:** A doesn't know that B has a car.
   a. A: Why are you mad?
   b. B: My car got stolen.

In order to explain cases like (1), proponents of the satisfaction theory assume that there is a process of *Global Accommodation*, by which B can accommodate a common ground which does satisfy the presupposition and then evaluate the assertion with respect to that common ground. In (1), A therefore has to accommodate a common ground that entails that B has a car before evaluating the assertion with respect to the accommodated common ground.

The availability of global accommodation raises the question of whether we can ever detect a difference between the pragmatic status of presuppositions and assertions. In particular, what is the distinction between accommodating a novel common ground in order to satisfy a presupposition and updating the common ground by accepting an assertion [1]? If assertions are evaluated with respect to the accommodated common ground, then conditions on assertability have to be evaluated after accommodation. In principle, this should allow us to detect an asymmetry between presupposed content which has to be accommodated, and asserted content.

In the rest of the paper, we discuss a constraint on accommodation proposed by Heim, which relies on this asymmetry between accommodation and assertion. In section 2, we present evidence for this constraint and show how it follows from the satisfaction theory. In section 3, we argue that this asymmetry between presupposition and assertion is maintained when

*Thanks to Amir Anvari, Ido Benbaji, Kai von Fintel, Danny Fox, Janek Guerrini, Martin Hackl, Adèle Mortier, Keely New, Lorenzo Pinton, Philippe Schlenker and Yash Sinha for their helpful feedback.

[1] See [AFH22] and the last section of [vF08] for a more detailed discussion of this puzzle.

presuppositions are locally accommodated. We discuss how this can be conceptualized under different theories of local accommodation.

2 Global accommodation

Heim argues that questions can't be answered with an accommodated presupposition [Hei15]. This constraint follows from the satisfaction view, once we make explicit certain conditions on assertability. In particular, in order for a proposition to be asserted it has to be informative, i.e. given a QUD which partitions the common ground, it has to eliminate some cells from the partition (2). The satisfaction view predicts that informativity has to be evaluated with respect to the common ground after accommodation, a requirement which can be formulated as shown in (3). If presupposition accommodation settles the QUD, then the assertion can't be informative after accommodation, violating the constraint in (3).

(2) A proposition p is informative with respect to a QUD Q and a common ground C **iff** $\exists q \in$ Q: $q \cap C \neq \emptyset \wedge p \cap q = \emptyset$

(3) **Post-Accommodation Informativity (PAI):** A sentence S_p (presupposing p) can be uttered felicitously only if S_p is informative w.r.t the QUD and common ground after presupposition accommodation.

Consider the example in (4) [Hei15]. After accommodating the presupposition that B has a sister, A's question is completely settled, making the assertion in (4-b) trivial with respect to the question. We can contrast this with (5), where accommodation is licensed. Here, accommodating that B has a sister does not settle the question, allowing the assertion to be informative with respect to the post-accommodation common ground [2].

(4) **Context:** A doesn't know B very well, and specifically knows nothing about her family.
   a. A: Do you have any siblings?
   b. B: # I need to pick up my sister from the airport.

(5) a. A: Are you coming to lunch?
   b. B: I need to pick up my sister from the airport.

The contrast between (4) and (5) shows that the QUD plays a crucial role, but it is not clear whether the infelicity of (4) is due to an asymmetry between presupposed and asserted content. In particular, it is possible that (4-b) is not a good answer because it provided additional information that is not relevant to the question.

In order to conclusively show the role of the asymmetry in this constraint, we have to look at minimal pairs which provide the same information but differ in terms of the division of labor between presupposition and assertion. Consider (6): note that after collapsing presupposition and assertion, (6-b) and (6-c) are completely equivalent – they are both true iff someone from NY adopted the Lab. Therefore, the infelicity of (6-c) has to be due to the presupposition that someone adopted the Labrador. Furthermore, given the availability of global accommodation, we can't attribute this simply to the fact that the presupposition is not met in the context. The infelicity therefore has to be explained by the constraint on accommodation in (3): accommodation is ruled out because the assertion that someone from NY adopted the Lab is uninformative with respect to the QUD after accommodating that someone adopted the Lab.

[2] It is sometimes possible for the hearer to accommodate an alternative QUD which satisfies PAI. The examples in this section try to control for this by making accommodation of such a QUD implausible.

(6) **Context:** A is visiting a dog shelter and is interested in adopting a Lab that she saw the week before when passing by the shelter. She comes back and asks an employee:
   a. A: Can I adopt the Lab?
   b. B: Someone from NY adopted the Lab.
   c. B': # It is someone from NY who adopted the Lab. [AFH22] (slightly modified)

Note that global accommodation with a cleft is in general licensed, given an appropriate QUD. In (7), the assertion that Mary adopted the Lab is still informative after accommodating that someone adopted the Lab, and A can therefore accommodate the presupposition. This provides further support that the infelicity of (6-c) is due to PAI and not due to a ban on accommodation with clefts in general.

(7) **Context:** A and B work at a dog shelter. B loves the Labrador and wants someone responsible to adopt it. A realizes that B is really upset. A also doesn't know that Mary, who they both know to be a careless person, adopted the Labrador the day before.
   a. A: Why are you upset?
   b. B: It is Mary who adopted the Lab.

Another example is given below. In (8-b), the assertion is trivial after accommodating that B has two children, which settles the question. In (9), we see that accommodation is licensed given an appropriate QUD. Finally, (8-c), under the non-restrictive reading of *adopted*, provides the same information as (8-b) if we collapse presupposition and assertion. This shows that it is the presupposition that causes the infelicity of (8-b).

(8) a. A: How many children do you have?
   b. B: # I adopted both of my children.
   c. B': I have two adopted children.

(9) **Context:** A and B are in a group for parents who have adopted children. A doesn't know B very well, and specifically knows nothing about her family.
   a. A: How many children have you adopted?
   b. B: I adopted both of my children.

We have therefore seen evidence supporting the view that informativity is evaluated after presupposition accommodation. This allowed us to diagnose an asymmetry between the presuppositional and assertive components of an utterance that follows from the satisfaction view of presuppositions. In the next section, we argue that local accommodation maintains this asymmetry, a result which is surprising given certain views of local accommodation.

3 Local accommodation

3.1 PAI for local accommodation

A characterizing property of presuppositions is that they project from certain embedded environments, such as negation and the antecedent of conditionals. But in some cases, this property of presuppositions disappears, as shown in (10) and (11). In both examples, not only is the presupposition of the embedded sentence not satisfied by the CG, it also contradicts the first sentence. Therefore, it is clear that no global accommodation is taking place. These facts have been explained by arguing that the presupposition of the embedded sentence is accommodated locally – under the negation in (10) and inside the antecedent in (11).

(10) The king of France didn't come. France doesn't have a king. [Hei83]

(11) I don't know whether John is married. But if John's wife came to the party, then he enjoyed it. [Pea10]

It's been claimed that some triggers allow local accommodation more easily than other triggers [Abu02, a.o.]. Therefore, in order to test whether PAI extends to local accommodation, we consider two *weak* triggers – *stop* and the possessive *her* – and focus our discussion on the antecedent of conditional sentences, leaving other environments for future work.

Consider (12): the only difference between (12-a) and (12-b) is that (12-a) presupposes that it was raining earlier. Why can't the presupposition of *stop* be locally accommodated in the antecedent in (12-a)? We argue that this is because accommodating the presupposition (that it was raining earlier) makes the antecedent trivial[3], given that it is common ground that it isn't raining now. Note that in (13) where it is not common ground that it isn't raining, local accommodation is licensed with the same sentence.

(12) **Context:** We've just left the movie theater and are standing in the exit hall, discussing what to do next. The morning forecast said that it might rain while we're at the movies, but we can all clearly hear that it isn't raining now.

a. # I don't know if it rained, but if it stopped raining, we might see a rainbow.
b. I don't know if it rained, but if it did rain and stopped, we might see a rainbow.

(13) **Context:** We're at the movie theater still in our seats, talking about what to do when we get out. The morning forecast said that it might rain while we're at the movies.
I don't know if it rained, but if it stopped raining, we might be able to see a rainbow.

Another example that illustrates this constraint is given in (14). Local accommodation is not licensed in (14-a), where it's common ground that Mary drove her vehicle to the party, since accommodating that Mary has a car in (14-a) makes the antecedent trivial. On the other hand, given an alternative common ground in (15), local accommodation is licensed.

(14) **Context:** We know that Mary drove her vehicle to the party and that it's is either a car or a motorcycle. We're wondering how we can get back home from the party.

a. # I don't know if Mary has a car or if she has a motorcycle, but if she drove her car here, she can drop us off.
b. I don't know if Mary has a car or if she has a motorcycle, but if she has a car and drove it here, she can drop us off.

(15) **Context:** We don't know how Mary got to the party, and we want to get back home:
I don't know if Mary has a car, but if she drove her car here, she can drop us off.

The data in (12-15) motivates the generalization in (16), which states that local accommodation is not licensed if the antecedent is trivial after updating the CG with its presupposition.

(16) Accommodation in the antecedent of 'if p_r, q' is not licensed **if** $C \cap r \rightarrow p$

In section 2, we saw that in the global accommodation case, PAI follows from the satisfaction view of accommodation, coupled with independently motivated constraints on assertion. We might have therefore expected PAI to be a strictly global phenomenon. The fact that we see

[3]Note that in both (12-a) and (14-a), the assertion of the antecedent is trivial post-accommodation regardless of the QUD. We constructed the examples that way to avoid making controversial assumptions about the relation between conditionals and QUDs.

a similar asymmetry with local accommodation raises interesting questions, both about the nature of local accommodation and the conception behind PAI. In what follows, we consider two different conceptions of local accommodation and discuss what assumptions are needed under each approach to explain the facts we have seen in this section.

3.2 Two conceptions of local accommodation

The nature of local accommodation is not obvious, especially in the Stalnakerian view of global accommodation as a pragmatic phenomenon. One class of theories argues that local accommodation is essentially a version of global accommodation, which takes place at the sub-sentential level [Hei83, a.o.]. We call this class of accounts **Unified Accommodation** (UA) theories.

In principle, UA-theories maintain the asymmetry in the status of presupposed and asserted content in local contexts. Looking specifically at the antecedent of conditionals, the predictions depend on whether there's a local requirement that targets informativity in the antecedent. If there is such a constraint, PAI predicts that it should be evaluated post-accommodation.

Note that conditionals are infelicitous when the antecedent is trivial with respect to the common ground, as shown in (17)[4]. There is a question of whether this constraint is a local informativity requirement on the antecedent (18) or some global requirement. Assuming that it is local, we argue that UA-theories predict the data in section 3.1[5].

(17) #I know that Mary read the book. If she read the book, she passed the exam.

(18) **Antecedent Triviality:** The antecedent of a conditional must be informative with respect to the local context of the antecedent.

Given PAI, we expect that (18) has to be evaluated after the antecedent's presupposition is accommodated in the local context. We therefore predict that the antecedent should be informative with respect to the common ground after presupposition accommodation – i.e. that local accommodation is not licensed under the conditions given in (16).

We therefore see that UA-theories predict the constraint in (16), but only under the assumption that there is a local triviality requirement on the antecedent of conditionals. Looking at more environments, the predictions are that we find this asymmetry only in environments where there is a local triviality requirement. To make clear predictions, we therefore need a general theory of this type of requirement. Nevertheless the UA-theories allow us to maintain that PAI follows from an asymmetry between the pragmatic status of presuppositions and assertions.

A second line of theory sees local accommodation as completely separate from global accommodation. This class of accounts views local accommodation as the result of an operation that collapses the presuppositional and asserted content of a proposition, and may apply at any scope site [BK01]. This operation is traditionally realized as the *A-operator*, given in (19). We call this class of accounts **Operator (O)** theories.

(19) $[\![\text{A}]\!](\text{p}) = \begin{cases} 1 \text{ iff } p = 1 \\ 0 \text{ iff } p = \# \vee p = 0 \end{cases}$ [Boc81]

Under the O-theories, we don't necessarily expect locally-accommodated presuppositions to behave differently from assertions, since all the *A*-operator does is collapse the two. The facts we have presented indicate that this cannot be the entire story, otherwise even the global cases

[4]There are some special cases, such as deduction contexts, where an antecedent may be trivial. We set these aside for the purposes of this paper.

[5]See [BDar] for further evidence that informativity is enforced locally.

like (6-c) and (8-b) should have been rescued by inserting A at the matrix level. The fact that we can ever detect PAI at work means that A must be sensitive to similar considerations.

One way to cash this out is to assume that the distribution of A is constrained in a way that prevents it from applying to a proposition if its asserted content does not contribute anything to the meaning of the entire sentence. More formally, we define an operator Prs which takes a trivalent proposition and extracts its presupposition (20-a). We propose that the A operator can't be inserted at any position where replacing it with Prs results in an equivalent meaning with respect to the context and QUD. This is formulated in (20-b).

(20) a. $[\![\text{Prs}]\!](\text{p}) = \begin{cases} 1 \text{ iff } p = 1 \vee p = 0 \\ 0 \text{ iff } p = \# \end{cases}$

b. **Distribution of A:** Given a sentence Φ that contains Ψ, and given a CG C and QUD Q:
$\Phi[\Psi \rightarrow A[\Psi]]$ is infelicitous if $[\![\Phi[\Psi \rightarrow A[\Psi]]]\!] \Leftrightarrow_{Q,C} [\![\Phi[\Psi \rightarrow Prs[\Psi]]]\!]$, where:
(i) Given a structure Φ that contains a constituent η, $\Phi[\eta \rightarrow \theta]$ is the structure where η is replaced by θ.
(ii) Given two propositions p, r: $p \Leftrightarrow_{Q,C} r$ **iff**
$\{q \in Q : q \cap C \neq \emptyset \wedge q \cap p = \emptyset\} = \{q \in Q : q \cap C \neq \emptyset \wedge q \cap r = \emptyset\}]$

Notice that (20-b) gives us the global version of PAI as a special case in which $\Phi = \Psi$. In other words, if the assertion is uninformative after accommodation, then the the result of applying Prs is necessarily equivalent to the result of applying A. Let's consider the example with the cleft (6-c). Here the presupposition (That someone adopted the Labrador) and the resulting proposition after applying A (That someone from New York adopted the Labrador) are equivalent with respect to the QUD: they both eliminate the cell where A can adopt the Labrador. Therefore, we predict that A can't be inserted in (6-c) given the QUD in (6).

The same is true for the conditional antecedents. In particular, given a conditional 'if p_r, q' and a common ground C where p_r is trivial with respect to $C \cap r$, then $A(p_r)$ is necessarily contextually equivalent to r. Therefore, replacing $A(p_r)$ with r in the antecedent will result in an equivalent proposition. This can be demonstrated using the example in (12). Given that it's common ground that it isn't raining now, the presupposition of the antecedent (that it was raining before) is contextually equivalent (regardless of the QUD) to the resulting proposition after applying A (that it was raining before and no longer is). Therefore, replacing A with Prs in the antecedent yields an equivalent meaning, and A can't be inserted.

This account of the local accommodation data yields clear predictions regarding the licensing of A in any environment – it should only be licensed if the equivalence in (20-b) doesn't hold. Notice that while we get the correct predictions, the O-theories demand us to posit a completely stipulative distribution rule for A, which doesn't follow from any pragmatic considerations.

4 Conclusion

In this paper, we have presented a constraint on global accommodation which follows from the satisfaction theory of presuppositions coupled with an informativity constraint on utterances. We have shown that a similar constraint holds for local accommodation in the antecedent of conditionals as well. In UA-theories, this follows from assuming a local triviality constraint on antecedents of conditionals, while in O-theories, we're required to posit a constraint on the distribution of the A operator. In future work, we will test the different predictions of these two approaches for other environments, e.g. the scope of quantifiers.

References

[Abu02] Dorit Abusch. Lexical alternatives as a source of pragmatic presuppositions. In *Semantics and linguistic theory*, volume 12, pages 1–19, 2002.

[AFH22] Athulya Aravind, Danny Fox, and Martin Hackl. Principles of presupposition in development. *Linguistics and Philosophy*, pages 1–42, 2022.

[BDar] Ido Benbaji and Omri Doron. Adversative *only* is only *only*. In *Sinn und Bedeutung*, volume 27, To appear.

[BK01] David Beaver and Emiel Krahmer. A partial account of presupposition projection. *Journal of logic, language and information*, 10(2):147–182, 2001.

[Boc81] Dmitri A Bochvar. On a three-valued logical calculus and its application to the analysis of the paradoxes of the classical extended functional calculus. *History and Philosophy of Logic*, 2(1-2):87–112, 1981.

[vF08] Kai von Fintel. What is presupposition accommodation, again? *Philosophical perspectives*, 22:137–170, 2008.

[Hei83] Irene Heim. On the projection problem for presuppositions. *Formal semantics–the essential readings*, pages 249–260, 1983.

[Hei15] Irene Heim. Presupposition. class notes, 24.954 (MIT, Fox/Heim), 2015.

[Kar74] Lauri Karttunen. Presupposition and linguistic context. 1974.

[Pea10] Hazel Pearson. Presupposition accommodation in local contexts: why global accommodation is not enough. *Ms Harvard University*, 5, 2010.

[Sta70] Robert C. Stalnaker. Pragmatics. *Synthese*, 22(1/2):272–289, 1970.

An alternative semantics for presupposition*

Julian Grove

University of Rochester
julian.grove@gmail.com

Abstract

Charlow [3, 5, 4] develops a view of indefinites according to which they give rise to alternatives, an aspect of their meaning intended to explain the fact that they may generally take exceptional scope past scope islands. [5, 4] proposes that exceptional scope arises from the confluence of two factors: two polymorphic type shifts that regulate how the alternative sets denoted by indefinites compose with other expressions, and the availability of roll-up pied piping as a scope-taking strategy. The current paper explores a related observation about presupposition triggers: they, too, involve a kind of exceptional scope, which is reflected in their projection behavior. I propose a move analogous to that of [9] by treating definite noun phrases (and other presupposition triggers) on a par with indefinites, but within a variant of the alternative semantics introduced by Charlow. The resulting interpretation scheme preserves the alternative-style analysis of indefinites, but integrates a novel analysis of presupposition triggers that accounts for their flexible projection behavior. It additionally yields a simple analysis of presupposition accommodation according to which denotations for presupposition triggers are mapped onto those for indefinites.

1 The alternative view of exceptional scope

In recent work, Charlow [3, 5, 4] develops the view that indefinite noun phrases give rise to alternatives, and that this aspect of their meaning is responsible for their ability to take exceptional scope past the scope islands created by finite clause boundaries.

(1) If Theo has a brother, he'll bring a wetsuit that he owns.

(1) has two distinct readings. The first reading, which gives the indefinite wide scope, entails that Theo owns a wetsuit, and that he will bring it if he has a brother. The second reading, which gives the indefinite narrow scope, entails more weakly that Theo owns wetsuit if he has a brother. Charlow [5] analyzes such patterns as arising from the confluence of two factors: the indefinite's alternative-generating semantics and the availability of *roll-up pied piping* as a scope-taking strategy. According to the roll-up analysis, the indefinite first takes scope inside the consequent clause of the conditional; then, this clause itself takes scope above the conditional as a whole. This result is achieved by invoking two polymorphic type shifts, $\eta_{\mathtt{S}}$ and $\star_{\mathtt{S}}$, that can be interleaved in a semantic derivation with the rules of Functional Application and Predicate Abstraction [10]. These type shifts are defined in the following way, where '$\mathtt{S}\alpha$' abbreviates the type $\alpha \to t$ (it is the type of sets of alternative α's):

$$\eta_{\mathtt{S}} : \alpha \to \mathtt{S}\alpha \qquad (\star_{\mathtt{S}}) : \mathtt{S}\alpha \to (\alpha \to \mathtt{S}\beta) \to \mathtt{S}\beta$$

$$\eta_{\mathtt{S}}(v) = \{v\} \quad (= \lambda x.x = v) \qquad m \star_{\mathtt{S}} k = \bigcup_{x \in m} k(x) \quad (= \lambda y.\exists x : m(x) \wedge k(x)(y))$$

*Thanks to Dylan Bumford and Simon Charlow for helpful comments on the ideas presented here. Any errors are mine.

$$\begin{aligned} \eta(v) \star k &= k(v) && \text{(Left Identity)} \\ m \star \lambda x.\eta(x) &= m && \text{(Right Identity)} \\ (m \star n) \star o &= m \star \lambda x.(n(x) \star o) && \text{(Associativity)} \end{aligned}$$

Figure 1: The monad laws

Together with these components, $\mathtt{S}$, viewed as a type constructor, gives rise to a *monad*; that is, a map M from types to types associated with operators η_M and $\star_M$ having the type signatures

$$\begin{aligned} \eta_M &: \alpha \to M\alpha \\ (\star_M) &: M\alpha \to (\alpha \to M\beta) \to M\beta \end{aligned}$$

and satisfying the laws in Figure 1. Charlow exploits this fact. For example, the law of Associativity allows roll-up pied piping to generate exceptional scope, since it means that whole-clause scope-taking can simulate the effect of the indefinite having taken scope above the conditional on its own in (1) (see [5] for discussion).

These points are illustrated in Appendix A, Figure 2, where exceptional scope for (1) is achieved. Here, the function `if` is defined as $\lambda\phi,\psi.\{\phi^{\natural} \to \psi^{\natural}\}$; following [5], it uses an *evaluation* function $(\cdot)^{\natural}$ turning a set of alternative truth values into a truth value:

$$\begin{aligned} (\cdot)^{\natural} &: \mathtt{S}t \to t \\ \phi^{\natural} &= \exists p : p \in \phi \wedge p \end{aligned}$$

`if` thus has the semantics of a material conditional while also evaluating its two arguments, effectively by checking that each of them returns $\top$ ('true') as once of its alternatives. Finally, the result of the derivation in Figure 2 is a set of alternative truth values; it evaluates via $(\cdot)^{\natural}$ to $\top$ just in case Theo owns a wetsuit which, if he has a brother, he'll bring.

The above is meant to give a gist of the workings of Charlow's system. The main goal of this paper is to slightly recast the alternative-monadic composition scheme, in order to countenance the semantic definedness conditions generated by presupposition triggers, along with the alternatives generated by indefinites. This is accomplished essentially by enriching the space inhabited by the functions characterizing alternative sets. While one could mix an analysis of presupposition triggers into an alternative semantics for indefinites in more than one way, I believe two considerations motivate the current move. First, as briefly discussed in the next section, presupposition triggers display exceptional scope behavior mirroring that displayed by indefinites; when presupposition triggers take scope, it is manifested not as non-determinism (as it is for indefinites), but as presupposition projection. This at least suggests that a monad should be available to countenance both.[1] Second, presupposition triggers, I argue, sometimes show up in a guise indicating that they should be analyzed as indefinites; that is, when they are locally accommodated [8]. In §4, it is shown that an alternative-style analysis of presupposition triggers makes a semantic analysis of presupposition accommodation straightforward to state.

[1] Such a monad could, in principle, arise from the application of a monad transformer to $\mathtt{S}$, for example.

2 The exceptional scope of presupposition triggers

Remarkably, presupposition triggers appear to give rise to a pattern of projection behavior similar to the scopal behavior displayed by indefinites.

(2) If Theo has a brother, he'll bring his wetsuit.

(2) is most easily understood as presupposing that Theo has a wetsuit; but it may also be understood with the weaker presupposition *if Theo has a brother, he has a wetsuit*, given appropriate background information. This is illustrated in (3) (taken from [7]).

(3) We're going scuba diving later, and I don't know if Theo owns a wetsuit. But it seems that everyone who has a brother got a wetsuit for Christmas. So, if Theo has a brother, he'll bring his wetsuit.

Following [7], I assume that sentences with presupposition triggers are in general semantically ambiguous, though I won't further argue this point here (see [7] for a more developed discussion).

Geurts [6] points out that examples like (2) provide problems for satisfaction accounts of presupposition projection in the vein of Heim [8], since such accounts predict them to give rise to conditional presuppositions instead of unconditional ones. As illustrated in [7], a monadic account of presupposition that percolates undefined values, when used in tandem with roll-up pied piping, may supplement satisfaction accounts to achieve unconditional presuppositions for such examples. This strategy of explanation assimilates presupposition projection to scope-taking; global projection of a presupposition results from the presupposition trigger taking wide scope, while local projection results from narrow scope.

In the next section, I show how to treat presupposition projection in terms of an alternative semantics. In these terms, an analysis of the ambiguity of (2) may be wholly assimilated to an analysis of the ambiguity of (1). Indeed, both (in principle) display either the conditional inference that Theo owns a wetsuit if he has a brother, or the unconditional inference that he owns a wetsuit. What differs between the analyses of each is whether the relevant inference is hypothesized to arise by evaluating an alternative set encoding a truth condition or one encoding a definedness condition.

3 Alternatives and definedness at once

Heim [9] treats definites and indefinites as essentially similar in their semantic contributions, differing in the familiarity/novelty requirements they impose on the common ground. The move I make here is somewhat analogous, but it is taken within a variant of the Charlow-style semantics recalled in §1. It thus preserves the alternative-style analysis of indefinites; meanwhile, it fully integrates an analysis of presupposition triggers that regards them as giving rise to semantic definedness conditions.

As a first step, I adopt a trivalent semantics for truth values by introducing a type $t_{\#}$, inhabited by $\top$, $\bot$, along with a third "undefined" value $\#$. The type t, which is inhabited by only $\top$ and $\bot$, is thus a subtype of $t_{\#}$.[2] It is also useful to define a new binary connective $\wedge_{\#}$, as well as a new existential quantifier $\exists_{\#} : (\alpha \to t_{\#}) \to t_{\#}$, having the following semantics:

[2]The '$\top$' and '$\bot$' notation, and that for expressions of type t more generally, will equivocate between values of type t and those of type $t_{\#}$, a convention justified by viewing t as a subtype of $t_{\#}$. Throughout, it will be clear from context (e.g., a type annotation) which type is intended.

$\wedge_\#$	$\top$	$\bot$	$\#$
$\top$	$\top$	$\bot$	$\#$
$\bot$	$\bot$	$\bot$	$\#$
$\#$	$\#$	$\#$	$\#$

$$\exists_\# x : \phi(x) = \begin{cases} \top & \exists x : (\phi(x) = \top) \\ \bot & \exists x : (\phi(x) = \bot) \wedge \neg\exists x : (\phi(x) = \top) \\ \# & \forall x : (\phi(x) = \#) \end{cases}$$

The connective $\wedge_\#$ effectively has a weak Kleene semantics, percolating up undefinedness from either of its conjuncts. Meanwhile, $\exists_\#$ is highly charitable as an existential quantifier, resulting in an undefined value only if its scope is undefined for all possible values of the quantified variable. (Otherwise, it acts similar to a classical existential quantifier.) Finally, it is useful to introduce an operator $\delta : t \to t_\#$ akin to the δ-operator of [2]; that is, $\delta(\top) = \top$ and $\delta(\bot) = \#$.

Given these pieces, one may redefine the operators from the first section, including the evaluation function, with the new type $t_\#$ in place of t. Characterizing sets as functions of type $\alpha \to t_\#$ instead of functions of type $\alpha \to t$ means that membership in an alternative "set" may be *undefined*; that is, where the value of its characteristic function is #. Here, '$\mathtt{S}_\#\alpha$' abbreviates the type $\alpha \to t_\#$.

$$\eta_{\mathtt{S}_\#} : \alpha \to \mathtt{S}_\#\alpha \qquad (\star_{\mathtt{S}_\#}) : \mathtt{S}_\#\alpha \to (\alpha \to \mathtt{S}_\#\beta) \to \mathtt{S}_\#\beta \qquad (\cdot)^{\downarrow\#} : \mathtt{S}_\# t_\# \to t_\#$$
$$\eta_{\mathtt{S}_\#}(v) = \lambda x.\delta(x = v) \qquad m \star_{\mathtt{S}_\#} k = \lambda y.\exists_\# x : m(x) \wedge_\# k(x)(y) \qquad \phi^{\downarrow\#} = \exists_\# p : \phi(p) \wedge_\# p$$

This encoding of alternatives also gives rise to a monad (Theorem 2 of Appendix C). As a consequence, the exceptional scope behavior characteristic of monadic semantics carries over.

Using $\mathtt{S}_\#$ in place of $\mathtt{S}$ allows for the derivation of (2) in Figure 5 of Appendix B. Here, *his wetsuit* is taken to denote $\lambda x.\delta(\mathsf{suitOfT}(x))$ — that is, the characteristic function of a "set" whose value is $\top$ on Theo's wetsuit (if it exists) and # everywhere else. The meaning of *if* in this derivation ($\mathtt{if}_\#$) is taken to be the function $\lambda\phi, \psi.\eta_{\mathtt{S}_\#}(\phi^{\downarrow\#} \Rightarrow \psi^{\downarrow\#})$ of type $\mathtt{S}_\# t_\# \to \mathtt{S}_\# t_\# \to \mathtt{S}_\# t_\#$. The connective $\Rightarrow$ featured here takes two values of type $t_\#$ onto a third and is assumed to have a semantics in the style of [12]:

$\Rightarrow$	$\top$	$\bot$	$\#$
$\top$	$\top$	$\bot$	$\#$
$\bot$	$\top$	$\top$	$\top$
$\#$	$\#$	$\#$	$\#$

Given either $\top$ or $\bot$ as arguments, $\Rightarrow$ behaves as a material conditional. Otherwise, it automatically projects undefinedness in its left conjunct, while undefinedness in its right conjunct is projected only if the left conjunct is $\top$.

Analogous to the derivation of (1), the type resulting for (2) is $\mathtt{S}_\# t_\#$: that of a "set" of truth values (here of type $t_\#$), but, in this case, whose membership relation may not always be defined. If this set is evaluated using $(\cdot)^{\downarrow\#}$, a value of type $t_\#$ is produced. The representation of this value is quite complex, given the term decorating the root in Figure 5, but it may be

simplified by relying on the theorem δ-elim of Appendix C:

$$
\begin{aligned}
& (\lambda p.\exists_{\#}\phi, x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(\phi = \eta_{\mathsf{S}_{\#}}(\mathsf{bring}(x)(\mathsf{t}))) \wedge_{\#} \delta(p = ((\eta_{\mathsf{S}_{\#}}(\mathsf{bro}(\mathsf{t})))^{\natural\#} \Rightarrow \phi^{\natural\#})))^{\natural\#} \\
= \; & (\lambda p.\exists_{\#}\phi, x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(\phi = \eta_{\mathsf{S}_{\#}}(\mathsf{bring}(x)(\mathsf{t}))) \wedge_{\#} \delta(p = ((\exists_{\#}q : \delta(q = \mathsf{bro}(\mathsf{t})) \wedge_{\#} q) \Rightarrow \phi^{\natural\#})))^{\natural\#} \\
& \qquad \text{(by defs. of } (\cdot)^{\natural\#} \text{ and } \eta_{\mathsf{S}_{\#}}) \\
= \; & (\lambda p.\exists_{\#}x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(p = (\mathsf{bro}(\mathsf{t}) \Rightarrow (\eta_{\mathsf{S}_{\#}}(\mathsf{bring}(x)(\mathsf{t})))^{\natural\#})))^{\natural\#} \qquad \text{(by } \delta\text{-elim)} \\
= \; & (\lambda p.\exists_{\#}x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(p = (\mathsf{bro}(\mathsf{t}) \Rightarrow \exists_{\#}q : \delta(q = \mathsf{bring}(x)(\mathsf{t})) \wedge_{\#} q)))^{\natural\#} \\
& \qquad \text{(by defs. of } (\cdot)^{\natural\#} \text{ and } \eta_{\mathsf{S}_{\#}}) \\
= \; & (\lambda p.\exists_{\#}x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(p = (\mathsf{bro}(\mathsf{t}) \Rightarrow \mathsf{bring}(x)(\mathsf{t}))))^{\natural\#} \qquad \text{(by } \delta\text{-elim)} \\
= \; & \exists_{\#}p, x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} \delta(p = (\mathsf{bro}(\mathsf{t}) \Rightarrow \mathsf{bring}(x)(\mathsf{t}))) \wedge_{\#} p \qquad \text{(by def. of } (\cdot)^{\natural\#}) \\
= \; & \exists_{\#}x : \delta(\mathsf{suitOfT}(x)) \wedge_{\#} (\mathsf{bro}(\mathsf{t}) \Rightarrow \mathsf{bring}(x)(\mathsf{t})) \qquad \text{(by } \delta\text{-elim)}
\end{aligned}
$$

The semantics of $\exists_{\#}$ dictates that the value that results here is equal to # if Theo has no wetsuit, and to $\top$ or $\bot$ otherwise, as according to whether or not he will bring his wetsuit if he has a brother. Thus the presupposition predicted for (2) is that Theo has a wetsuit. Though it isn't provided here, the derivation of the conditional presupposition that Theo has a wetsuit *if he has a brother* can be generated for (2) by giving the presupposition trigger narrow scope — that is, by invoking $\eta_{\mathsf{S}_{\#}}$ one fewer time in the scope of the moved trigger and evaluating the consequent of the conditional *in situ*.

Meanwhile, from a monadic bird's-eye view, the analysis of (1) is almost unchanged. Figure 4 of Appendix B gives a derivation for (1) using $\mathsf{S}_{\#}$ which is equivalent to that given in Figure 2, save the details that (1) the monadic operators have been redefined, and (2) formulae that were of type t before are now of type $t_{\#}$. Evaluating the result with $(\cdot)^{\natural\#}$ yields a value whose representation can be simplified, as above, to yield:

$$\exists_{\#}x : \mathsf{suitOfT}(x) \wedge_{\#} (\mathsf{bro}(\mathsf{t}) \Rightarrow \mathsf{bring}(x)(\mathsf{t}))$$

The value of this formula is either $\top$ or $\bot$: $\top$ if Theo has a wetsuit that he will bring if he has a brother, and $\bot$ if not. What was a definedness condition in the analysis of (2) is a here a truth condition.

4 Accommodation

The combined alternative semantics treats indefiniteness and presupposition uniformly, so that indefinite expressions and presupposition triggers are of the same semantic types $\mathsf{S}_{\#}\alpha$ (for some α) and thus denote in the same space of functions into $t_{\#}$. The only semantic difference between the two kinds of expression, from the current perspective, is the way in which they use this space. Indefinites encode alternatives in terms of the values $\top$ and $\bot$: a live alternative is mapped to $\top$, while any non-alternative is mapped to $\bot$. Meanwhile, presupposition triggers encode presuppositions in terms of the values $\top$ and #: a value that *satisfies* an expression's presupposition is mapped to $\top$, while any value which doesn't is mapped to #. When values of type $\mathsf{S}_{\#}t_{\#}$ are evaluated via $(\cdot)^{\natural\#}$, these encodings are cashed out as existential truth conditions for indefiniteness and existential definedness conditions for presupposition triggers.

Treating the two as akin allows for a straightforward implementation of the phenomenon known as "local accommodation" [8], a species of the accommodation process described by

Lewis [11], but which appears inside the scope of linguistic operators, e.g., negation. We can look at Heim's original example of the phenomenon to illustrate it.

(4) The king of France didn't come.

Consider (as Heim does) a scenario in which (4) is uttered and then followed up with *because France doesn't have a king*. In such a context, (4) may be understood as denying that France has a king who came. In the spirit of [3], negation may be analyzed in terms of an operator `not` which cancels the alternatives generated by indefinites:

$$\begin{aligned} \texttt{not} &: \mathsf{S}_{\#} t_{\#} \to t_{\#} \\ \texttt{not}(\phi) &= \neg(\phi^{\mathsf{I}\#}) \end{aligned} \qquad \text{where} \qquad \begin{aligned} \neg\top &= \bot \\ \neg\bot &= \top \\ \neg\# &= \# \end{aligned}$$

Note that the $\neg$ connective leaves # unmodified. As a result, `not` will allow presuppositions (though not alternatives) to project past it. Finally, we may introduce a polymorphic operator `accom` which has the effect of turning definedness conditions into truth conditions:

$$\begin{aligned} \texttt{accom} &: \mathsf{S}_{\#}\alpha \to \mathsf{S}_{\#}\alpha \\ \texttt{accom}(m) &= \lambda x.\delta^{-1}(m(x)) \end{aligned} \qquad \text{where} \qquad \begin{aligned} \delta^{-1}(\top) &= \top \\ \delta^{-1}(\bot) &= \bot \\ \delta^{-1}(\#) &= \bot \end{aligned}$$

`accom` relies on an operator δ^{-1}, which is effectively a left inverse of δ: it leaves $\top$ and $\bot$ unmodified, but takes # back onto $\bot$.[3]

Though a lack of space precludes a full derivation, we may use these components to provide a meaning representation for (4) which is suggestive of one:

$$\begin{aligned} & \texttt{not}(\texttt{accom}(\lambda p.\exists_{\#} x : \delta(\mathsf{kof}(x)) \wedge_{\#} \delta(p = \mathsf{come}(x)))) \\ = \ & \texttt{not}(\lambda p.\delta^{-1}(\exists_{\#} x : \delta(\mathsf{kof}(x)) \wedge_{\#} \delta(p = \mathsf{come}(x)))) && \text{(by def. of } \texttt{accom}) \\ = \ & \neg(\exists_{\#} p : (\delta^{-1}(\exists_{\#} x : \delta(\mathsf{kof}(x)) \wedge_{\#} \delta(p = \mathsf{come}(x))) \wedge_{\#} p) && \text{(by def. of } \texttt{not}) \end{aligned}$$

By attending to the possible values of the formula serving as the argument to δ^{-1} here (and with the help of δ-elim), it is possible to see that this formula is equivalent to the following one:

$$\neg\exists_{\#} x : \mathsf{kof}(x) \wedge_{\#} \mathsf{come}(x)$$

Thus by first locally accommodating the presupposition of (4) via `accom`, we have managed to cancel it via `not`.

5 Conclusion

An alternative semantics encoding presuppositional definedness conditions allows an explanation of the semantic kinship between indefinites and definites (and presupposition triggers, generally) manifest in their similar scopal behavior and when presupposition triggers are accommodated. On the alternative analysis of presupposition, both give rise to some kind of existence inference: indefinites require the existence of an alternative (where their corresponding inference is *at issue*), while presupposition triggers have a similar requirement, though the stakes are about definedness rather than truth. Future work will have to determine how plausibly these assumptions can be extended to the variety of presupposition triggers that exist.

[3] It is thus akin to the A operator of [1].

A Exceptional scope via S

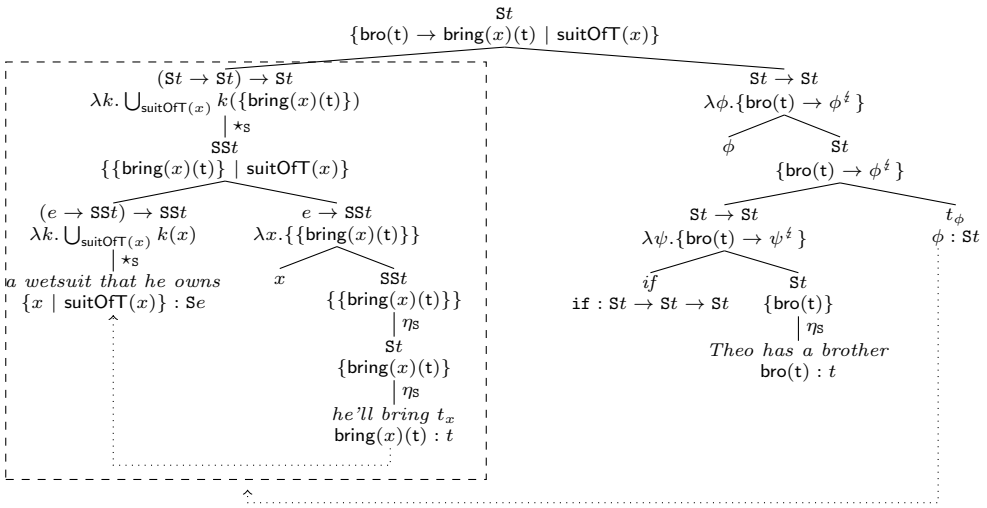


Figure 2: Deriving (1) via S

B Exceptional scope via S#

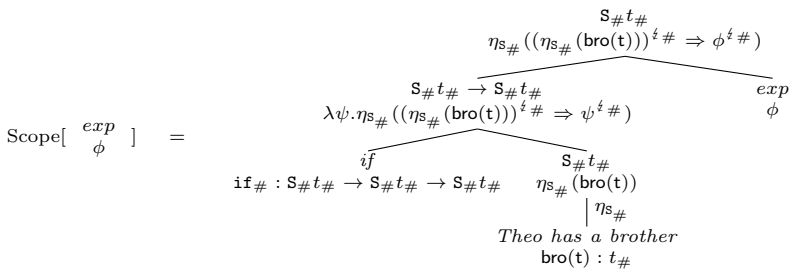


Figure 3: Scope of the moved expression in Figures 4 and 5

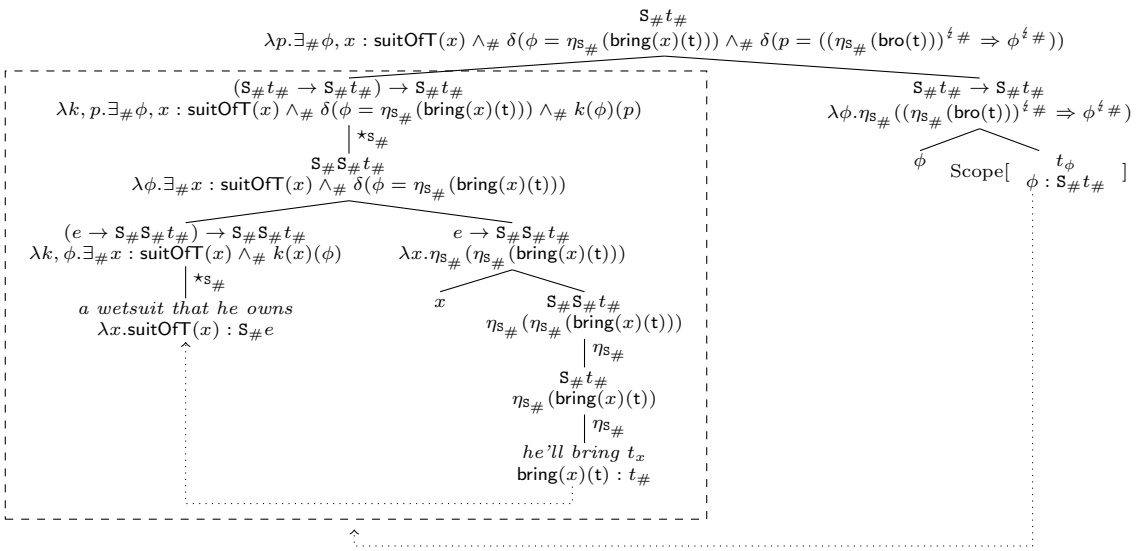


Figure 4: Deriving (1) via $\mathtt{S}_{\#}$

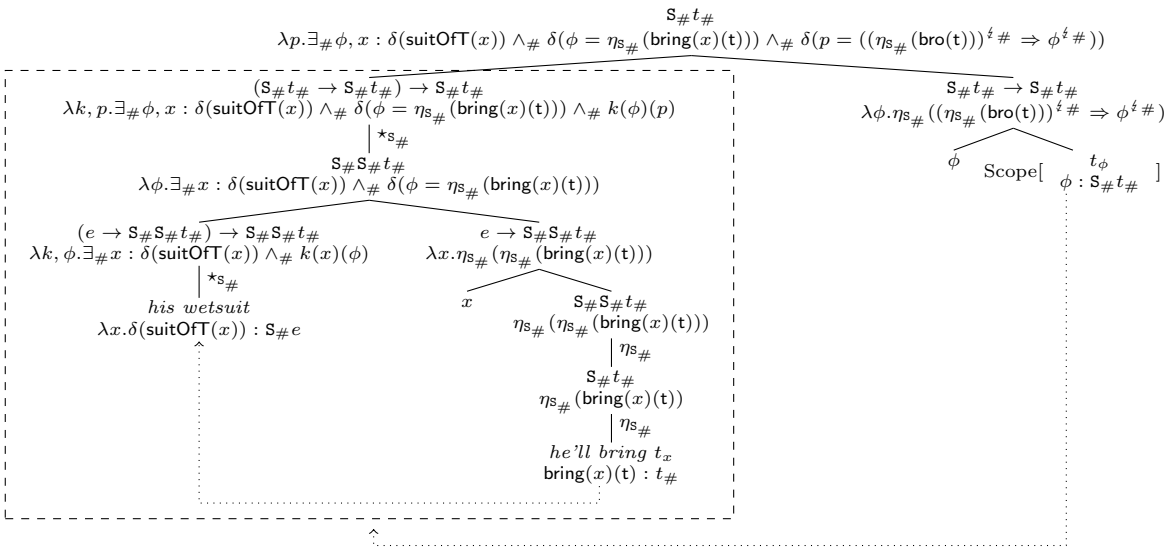


Figure 5: Deriving (2) via $\mathtt{S}_{\#}$

C Proof that $\mathtt{S}_{\#}$ is a monad

This section offers only proof sketches. The actual proofs are available at http://github.com/juliangrove/trivalent-alternatives.

The following theorem is useful for simplifying formulae that result from the composition of monadic operators.

Theorem 1 (δ-elim)**.** $\exists_{\#}x : \delta(x = v) \wedge_{\#} \phi(x) \quad = \quad \phi(v)$.

Proof. Both the LHS and RHS are either true, false, or undefined. If the LHS is true, then there is a witness to the existential quantifier which, when chosen as the value of x, makes its scope true; this witness must be v, making the RHS true. If the LHS is false, then choosing v as the value of x makes the quantifier's scope false, in turn making the RHS false. If the LHS is undefined, then choosing v as the value of x makes the quantifier's scope undefined, in turn making the RHS undefined. □

Theorem 2. $\mathtt{S}_{\#}$ is a monad.

Proof.
Left Identity

$$\begin{aligned}
\eta_{\mathtt{S}_{\#}}(v) \star_{\mathtt{S}_{\#}} k &= \lambda y.\exists_{\#}x : \delta(x = v) \wedge k(x)(y) && \text{(by def., } \equiv_{\beta}\text{)} \\
&= \lambda y.k(v)(y) && \text{(by } \delta\text{-elim)} \\
&= k(v) && \text{(by } \equiv_{\eta}\text{)}
\end{aligned}$$

Right Identity
This relies on $\wedge_{\#}$ being commutative.

$$\begin{aligned}
m \star_{\mathtt{S}_{\#}} \lambda x.\eta_{\mathtt{S}_{\#}}(x) &= \lambda y.\exists_{\#}x : m(x) \wedge_{\#} \delta(y = x) && \text{(by def., } \equiv_{\beta}\text{)} \\
&= \lambda y.m(y) && \text{(by commutativity of } \wedge_{\#}\text{, } \delta\text{-elim)} \\
&= m && \text{(by } \equiv_{\eta}\text{)}
\end{aligned}$$

Associativity
This relies on $\exists_{\#}$ commuting past $\wedge_{\#}$ when it binds a variable in one conjunt, $\exists_{\#}$ commuting with itself, and the associativity of $\wedge_{\#}$.

$$\begin{aligned}
(m \star_{\mathtt{S}_{\#}} n) \star_{\mathtt{S}_{\#}} o &= \lambda z.\exists_{\#}y : (m \star_{\mathtt{S}_{\#}} n)(y) \wedge_{\#} o(y)(z) && \text{(by def.)} \\
&= \lambda z.\exists_{\#}y : (\exists_{\#}x : m(x) \wedge_{\#} n(x)(y)) \wedge_{\#} o(y)(z) && \text{(by def., } \equiv_{\beta}\text{)} \\
&= \lambda z.\exists_{\#}y : \exists_{\#}x : (m(x) \wedge_{\#} n(x)(y)) \wedge_{\#} o(y)(z) && \text{(by comm. of } \exists_{\#}/\wedge_{\#}\text{)} \\
&= \lambda z.\exists_{\#}x : \exists_{\#}y : (m(x) \wedge_{\#} n(x)(y)) \wedge_{\#} o(y)(z) && \text{(by commutativity of } \exists_{\#}\text{)} \\
&= \lambda z.\exists_{\#}x : \exists_{\#}y : m(x) \wedge_{\#} (n(x)(y) \wedge_{\#} o(y)(z)) && \text{(by associativity of } \wedge_{\#}\text{)} \\
&= \lambda z.\exists_{\#}x : m(x) \wedge_{\#} \exists_{\#}y : n(x)(y) \wedge_{\#} o(y)(z) && \text{(by comm. of } \exists_{\#}/\wedge_{\#}\text{)} \\
&= \lambda z.\exists_{\#}x : m(x) \wedge_{\#} (n(x) \star_{\mathtt{S}_{\#}} o)(z) && \text{(by } \equiv_{\beta}\text{, def.)} \\
&= m \star_{\mathtt{S}_{\#}} \lambda x.(n(x) \star_{\mathtt{S}_{\#}} o) && \text{(by } \equiv_{\beta}\text{, def.)}
\end{aligned}$$

□

References

[1] David Beaver and Emiel Krahmer. A Partial Account of Presupposition Projection. *Journal of Logic, Language and Information*, 10(2):147–182, 2001. URL: https://doi.org/10.1023/A:1008371413822.

[2] David I. Beaver. The kinematics of presupposition. In Paul Dekker and Martin Stokhof, editors, *Proceedings of the Eighth Amsterdam Colloquium*, pages 17–36, Amsterdam. ILLC, 1992.

[3] Simon Charlow. *On the semantics of exceptional scope*. PhD Thesis, New York University, New York, 2014. URL: https://semanticsarchive.net/Archive/2JmMWRjY.

[4] Simon Charlow. Static and dynamic exceptional scope. Publisher: Rutgers University Published: LingBuzz, 2020. URL: https://ling.auf.net/lingbuzz/004650.

[5] Simon Charlow. The scope of alternatives: indefiniteness and islands. en. *Linguistics and Philosophy*, 43(4):427–472, 2020. ISSN: 1573-0549. DOI: 10.1007/s10988-019-09278-3.

[6] Bart Geurts. Local Satisfaction Guaranteed: A Presupposition Theory and Its Problems. *Linguistics and Philosophy*, 19(3):259–294, 1996. DOI: 10.1007/BF00628201.

[7] Julian Grove. Presupposition projection as a scope phenomenon. en. *Semantics and Pragmatics*, 15(15), December 2022. ISSN: 1937-8912. DOI: 10.3765/sp.15.15.

[8] Irene Heim. On the Projection Problem for Presuppositions. In Michael D. Barlow, Daniel P. Flickinger, and Michael Westcoat, editors, *The West Coast Conference on Formal Linguistics (WCCFL)*, volume 2, pages 114–125, Stanford. Stanford University Press, 1983.

[9] Irene Heim. *The Semantics of Definite and Indefinite Noun Phrases*. PhD Thesis, University of Massachussetts, Amherst, 1982. URL: https://semanticsarchive.net/Archive/Tk0ZmYyY/.

[10] Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, Malden, MA, 1998.

[11] David Lewis. Scorekeeping in a Language Game. In Rainer Bäurele and Arnim von Stechow, editors, *Semantics from Different Points of View*. Volume 6, Springer Series in Language and Communication, pages 172–187. Springer, Berlin, 1979. URL: https://doi.org/10.1007/978-3-642-67458-7_12.

[12] Stanley Peters. A truth-conditional formulation of Karttunen's account of presupposition. *Synthese*, 40(2):301–316, 1979. DOI: 10.1007/BF00485682.

Presupposition projection from the scope of quantifiers: A novel approach*

Alexandros Kalomoiros

University of Pennsylvania, Philadephia, PA, USA
akalom@sas.upenn.edu

Abstract

We present a modification of the Transparency Theory [5, 6], in the case of presupposition projection from the scope quantifiers, which allows the force of the quantifier to affect the force of the projected presupposition. The modification is based on letting the Transparency constraint operate on *witness sets* [1], and produces predictions that match currently known empirical results [3, 8], while also making novel predictions that will hopefully serve as input to future psycholinguistic experimentation.

1 Introduction

How presuppositions project from the scope of quantifiers is a vexed issue. Heim's Dynamic semantics, [4], predicted universal inferences across the board (the '▷' symbol below is to be read as 'presupposes'):

(1) Some/Every/At least n/No/At most n/... students stopped smoking. ▷ Every student used to smoke.

On the other hand [2] developed a theory where projection is existential across the board:

(2) Some/Every/At least n/No/At most n/... students stopped smoking. ▷ Some student used to smoke.

More recently, experimental forays paint a picture where the projected presupposition has universal force when the quantifier is universal, and existential force when the quantifier is existential, [8]. Importantly, the quantifier 'No' appears to behave like a universal, [3]. In cases where the quantifier is neither a full existential nor a full universal, the projected inference doesn't have robust universal force (see [3] for details).

Our aim in this paper is to see how the intuitions behind the Transparency Theory [5, 6] can be modified to produce results that match the aforementioned empirical picture. We argue that this can be achieved by utilising the notion of a *witness set* of a quantifier, leading to the following result:

(3) a. If $D^{\uparrow}$ is an upwards monotone determiner, then $D^{\uparrow}(students)(stopped\ smoking) \triangleright D^{\uparrow}(students)(used\ to\ smoke)$
b. If $D^{\downarrow}$ is a downwards monotone determiner, then $D^{\downarrow}(students)(stopped\ smoking) \triangleright D^{\downarrow}(students)(\neg(used\ to\ smoke))$

The rest of the paper is organised as follows: Section 2 introduces the necessary background on the Transparency Theory and witness sets. Section 3 presents our system and applies it to derive the result in (3). Section 4 concludes.

*Thanks to Florian Schwarz, Jacopo Romoli, Julie Legate, and to the members of the Penn semantics lab for their feedback at various stages of this project. All errors are my own.

2 Background

2.1 The Transparency theory

We give a brief description of the basics of the Transparency theory, [5, 6], to the extent needed to develop our idea about quantifiers. First, suppose we have a propositional language with two kinds of propositional constants p and $p'p$. p is a simple proposition denoting a set of worlds. $p'p$ is to be understood as the conjunction of p' and p, the idea being that p' is the presuppositional component, while p the non-presuppositional component. Assume also that all sentences are interpreted against a context C, modelled as a set of possible worlds [7].

Transparency Theory imposes a constraint on expressions of the form $p'p$ by requiring the presuppositional part to be uninformative/redundant in $p'p$ (with respect to the words in C). Here's the statement of the constraint:

(4) **Transparency:** A sentence S starting with a string $\alpha\ p'p$ is acceptable in a context C if and only if for all r, for all β it holds that $C \models \alpha\ p'r\ \beta \leftrightarrow \alpha\ r\ \beta$

What this constraint is requiring is that as soon as a presuppositional $p'p$ constituent is encountered in a sentence S, we are able to know that in every world in the context C, no matter how S ends (for all β), and no matter the non-presuppositional component of $p'p$ (for all r), S is equivalent to the version where the presupposition has been removed; the presupposition adds nothing, no matter the assertion and no matter the conclusion of the sentence.

To get a sense of how this works, consider the simple case of $\neg p'p$. The constraint requires that for all r:

(5) $C \models \neg p'r \leftrightarrow \neg r$

We can calculate a condition on the context that is equivalent to the constraint. First, suppose that the constraint holds. Since it holds for all r, take r to be some tautology $\top$. Then $\neg p'r$ becomes $\neg p'$ (recall $p'r$ is a conjunction) and the constraint becomes:

(6) $C \models \neg p' \leftrightarrow \neg\top$

The negation of $\top$ is a contradiction, so the right-hand side is false. Since we have assume the bi-conditional to be true, the left-hand side must also be false. For this to be the case, it must hold throughout C that p' is true.

Now suppose that p' is true throughout C. This is enough to make the constraint in (5) hold. If the left-hand side $\neg p'r$ is true in some world in C for some arbitrary r, then it must be that r is false in that world (since we assumed that p' is true in every world in C). But then the right-hand side is also true. If the right-hand side is true in some world for some arbitrary r, then $\neg p'r$ is also true. So, the equivalence holds just in case $C \models p'$, i.e. the presupposition p' 'projects' from the scope of the negation.

2.2 Projection from the scope of quantifiers

Now suppose we move to more complex language where we also have two types of (one-place) predicate symbols P and $P'P$. P is a normal predicate denoting a set of entities in some world i in C, while $P'P$ is a predicate that has P' as its presuppositional component, and denotes the intersection of the sets denoted by P' and P. We also have determiner symbols $D_j,\ j \in \mathbb{N}$, and quantificational sentences of the form $D_j(A)(B)$ which are interpreted in accordance with Generalized Quantifier theory, [1]. Following [5], we will assume that all of our determiners

obey Conservativity, Extension, and Permutation-invariance. We will refer to string of the form $D(A)$ as *quantifiers* (keeping to the definition of a generalized quantifier as a family of sets).

The case that interests us is that of a presuppositional expression in the scope of a quantifier, i.e. $D(A)(P'P)$. The constraint then becomes (see [5] for the exact details):

(7) For all R, it holds that $C \models D(A)(P'R) \leftrightarrow D(A)(R)$

[5] shows that under certain conditions, this constraint is equivalent to requiring that in every world $i \in C$, every entity in $[\![A]\!]^i$ be also in $[\![P']\!]^i$. Hence, a universal presupposition is projected whereby *Every/Some/Most/No/... student stopped smoking* presupposes that *Every student used to smoke*.

Our argument is that we can predict presuppositions that depend on the actual quantifier being used, if we shift the domain of application of the constraint to the *witness sets* that can be used to verify a quantificational sentence.

2.3 Witness sets

Barwise & Cooper, [1], famously suggested that when processing a quantificational statement of the form $D(A)(B)$, comprehenders are not necessarily calculating the entire set of properties denoted by $D(A)$ and then checking if $[\![B]\!]$ is in it. Instead they reason by using so-called *witness sets* of $D(A)$. Intuitively a witness set for a quantifier $D(A)$ is a subset of $[\![A]\!]$ that is of the correct quantificational amount for D. For instance, a witness set for $Most(stundets)$ is a subset of students that contains most of the students:

(8) a. $[\![students]\!] = \{a,\ b,\ c\}$
b. The set of witness sets for $Most(stundets)$ is $\{\ \{a,\ b,\ c\},\ \{a,\ b\},\ \{b,\ c\},\ \{c,\ a\}\ \}$

[1] define witness sets for quantifiers $D(A)$ that *live on* A, so first we give the definition of the *live on* property:

(9) A quantifier $D(A)$ **lives on** A iff for all X, $X \in [\![D(A)]\!] \leftrightarrow (X \cap [\![A]\!]) \in [\![D(A)]\!]$

For example $Some(man)$ lives on *man* since the following holds:

(10) [Some man VP] iff [Some man is a man who VP]

Note that all conservative determiners produce quantifiers that have the *live on* property: if D is conservative, then for all X, Y, $D(X)(Y) \leftrightarrow D(X)(X \cap Y)$. For $X = A$ then, it holds that for all Y, $D(A)(Y) \leftrightarrow D(A)(A \cap Y)$. Hence, when D is applied to some restrictor argument A it produces a quantifier that lives on A. Since we have restricted attention to determiners that satisfy Conservativity, all of our quantifier will have the *live on* property. With this out of the way, here's the definition of *witness set*:

(11) A **witness set** for a quantifier $D(A)$ living on A is any subset w of $[\![A]\!]$ such that $w \in [\![D(A)]\!]$

The reason [1] suggest that one use witness sets to reason about the truth or falsity of a quantificational expression $D(A)(B)$ is the following proposition:

(12) **Proposition 1** (Barwise and Cooper): Let w range over witness sets for a quantifier $D(A)$ living on A. Then, for any B:

- If D is upwards monotone, $D(A)(B)$ is true iff $\exists w[w \subseteq [\![B]\!]\ ]$
- If D is upwards monotone, $D(A)(B)$ is true iff $\exists w[\ [\![B]]\cap [\![A]\!] \subseteq w\ ]$

This proposition suggests a straightforward procedure for evaluating the truth or falsity of a quantificational statement $D(A)(B)$. If D is upwards monotone, then find a witness set w that is a subset of the property denoted by B. If D is downwards monotone, then find a witness set w of $D(A)$ such that $[\![A]\!] \cap [\![B]\!]$ is a subset of w. If no witness set of the right kind can be found, the sentence is false.

We are going to say that a witness set for a qunatifier $D(A)$ **verifies** a quantificational expression $D(A)(B)$ iff w fulfills the relevant condition from Proposition 1 above (depending on the monotonicity of D).

3 The current framework

We now combine witness sets with Transparency. Recall that in the case of $D(A)(P'P)$, the original Transparency constraint was:

(13) $\forall P:\ C \models D(A)(P'P) \leftrightarrow D(A)(P)$

Recall also the intuition behind this definition was that a presupposition should not contribute any information: when evaluating the truth of a presuppositional expression (no matter the assertion), this should be equivalent to ignoring the presupposition.

Our twist is the following: instead of requiring that P' be uninformative with respect to $D(A)(P'P)$, we are going to require that P' be uninformative with respect to a witness set that makes $D(A)(P'P)$ true:

(14) **Witness Set Transparency:** Given a quantificational expression $D(A)(P'P)$, it must hold that: $\forall P:\ C \models \exists w\ \ [w\ verifies\ D(A)(P'P)\ \leftrightarrow\ w\ verifies\ D(A)(P)]$

Parallel to the original Transparency intuition, this is saying that when in some world $i \in C$ you are picking a witness set to verify a presuppositional quantificational expression (no matter the assertive component), this should be equivalent to using this witness set to verify the same quantificational expression without the presupposition.

Before moving on, a note on the object-language: it will be helpful below to have access to a predicate-level tautology ($P_\top$), so we assume it as part of our language. The interpretation of $P_\top$ in some world i is just the domain of entities in i.

We are now going to apply our **Witness Set Transparency** to the case of upwards and downwards monotone determiners, in order to isolate necessary and sufficient conditions under which the **Witness Set Transparency** constraint holds for such determiners.

3.1 Upwards

Suppose that D is an upwards monotone determiner. We claim that **Witness Set Transparency** holds for $D(A)(P'P)$ if and only if in every world in C, there is a witness set w for $D(A)$ such that w verifies $D(A)(P')$.

To see this first suppose that **Witness Set Transparency** holds for $D(A)(P'P)$. This means that for all P in every world in C, there exists a witness set w for $D(A)$ such that, w verifies $D(A)(P'P)$ iff w verifies $D(A)(P)$. Since this holds for all worlds in C, take some arbitrary world i.

Since D is upwards monotone, w verifies $D(A)(P'P)$ iff $w \subseteq [\![P]\!]^i$ (by Proposition 1). Since this verification holds for all P, it must hold for $P = P_\top$. Thus it holds:

(15) $w \subseteq [\![P'P_\top]\!]^i \leftrightarrow w \subseteq [\![P_\top]\!]^i$

Since $[\![P'P_\top]\!]^i = [\![P']\!]^i \cap [\![P_\top]\!]^i$, this is just equal to $[\![P']\!]^i$. Thus we have:

(16) $w \subseteq [\![P']\!]^i \leftrightarrow w \subseteq [\![P_\top]\!]^i$

Trivially, it holds that $w \subseteq [\![P_\top]\!]^i$ (since $P_\top$ is just the set of all entities in i). Since we are assuming the bi-conditional in (16) to be true, this means that $w \subseteq [\![P']\!]^i$ is true, and by Proposition 1, this means that $D(A)(P')$ is true in i. Since i was arbitrary, this reasoning goes through for every world in C.

For the other direction, assume that that $D(A)(P')$ holds in some arbitrary C-world i. We need to show that **Witness Set Transparency** holds. Given that $D(A)(P')$ holds in i, Proposition 1 tells us that there is a witness set w for $D(A)$ such that $w \subseteq [\![P']\!]^i$. Consider this w; we are going to show that **Witness Set Transparency** holds for it, i.e. that:

(17) $\forall P : w$ verifies $D(A)(P'P)$ in $i \leftrightarrow w$ verifies $D(A)(P)$ in i

Take an arbitrary P, and suppose that w verifies $D(A)(P'P)$ in i. Then $w \subseteq [\![P'P]\!]^i$. But then all the elements in w are also in $[\![P]\!]^i$, hence $w \subseteq [\![P]\!]^i$, and w verifies $D(A)(P)$ in i.

Now suppose that w verifies $D(A)(P)$ in i. Then $w \subseteq [\![P]\!]^i$. But by assumption $w \subseteq [\![P']\!]^i$. So, all the elements of w are elements of both $[\![P']\!]^i$ and $[\![P]\!]^i$m which means that $w \subseteq [\![P'P]\!]^i$. But then w verifies $D(A)(P'P)$ in i. Since i was arbitrary, this reasoning goes through for all worlds in C, which means that **Witness Set Transparency** holds.

3.2 Downwards

Suppose that D is a downwards monotone determiner, and that we have a presuppositional quantificational expression $D(A)(P'P)$. We claim that **Witness Set Transparency** holds if and only if in every C-world, there is a witness set w for $D(A)$ such that w verifies $D(A)(\neg P')$.

To see this, first suppose that **Witness Set Transparency** holds for $D(A)(P'P)$. This means that for every P and every world in C, there exists a witness set w for $D(A)$ such that w verifies $D(A)(P'P)$ iff w verifies $D(A)(P)$.

Consider an arbitrary world $i \in C$. Since D is downwards monotone, the relevant w in i verifies $D(A)(P'P)$ iff $([\![A]\!]^i \cap [\![P]\!]^i) \subseteq w$ (by Proposition 1). Since this verification holds for all P, it must hold for $P = \neg P'$. Thus it holds:

(18) $([\![A]\!]^i \cap [\![P'(\neg P')]\!]^i) \subseteq w \leftrightarrow ([\![A]\!]^i \cap [\![\neg P']\!]^i) \subseteq w$

$[\![P'(\neg P')]\!]^i$ equals the empty set. Hence $([\![A]\!]^i \cap [\![P'(\neg P')]\!]^i)$ equals the empty set, and as the empty set is trivially a subset of any set, $([\![A]\!]^i \cap [\![P'(\neg P')]\!]^i) \subseteq w$ is true. Since we are assuming that the bi-conditional in (18) holds, this means that $([\![A]\!]^i \cap [\![\neg P']\!]^i) \subseteq w$ must hold. But then by Proposition 1, w verifies $D(A)(\neg P')$ in i. Since i was arbitrary, it holds throughout C that $D(A)(\neg P')$.

For the other direction, take some arbitrary C-world i and assume that there is w such that w verifies $D(A)(\neg P')$ in i. We need to show that **Witness Set Transparency** holds, i.e. that there is exists a witness set w' in i such that the following holds:

(19) $\forall P : w'$ verifies $D(A)(P'P)$ in $i \leftrightarrow w'$ verifies $D(A)(P)$ in i

Take an arbitrary P. Consider the witness set w such that w verifies $D(A)(\neg P')$ in i (given to us by assumption), and assume that w verifies $D(A)(P'P)$ in i. We argue that w also verifies $D(A)(P)$ in i. Since we assumed that w verifies $D(A)(P'P)$ in i, then $([\![A]\!]^i \cap [\![P'P]\!]^i) \subseteq w$. So w contains all the elements that are in $[\![A]\!]^i$ and $[\![P']\!]^i$ and $[\![P]\!]^i$. But by assumption it also holds that $([\![A]\!]^i \cap [\![\neg P']\!]^i) \subseteq w$, so w contains all the $[\![A]\!]^i$ elements that are in $[\![\neg P']\!]^i$. Therefore, the following holds:

(20) $([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i) \subseteq w$

We claim that $([\![A]\!]^i \cap [\![P]\!]^i) \subseteq ([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]^i) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i)$.

To see this, consider an element from $([\![A]\!]^i \cap [\![P]\!]^i)$. This element will be a member of either $[\![P']\!]^i$ or $[\![\neg P']\!]^i$. If it is an element of $[\![P']\!]^i$, then it is an element of $([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]^i)$ and hence is in $([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i)$. If it is an element of $[\![\neg P']\!]^i$, then it is in $([\![A]\!]^i \cap [\![\neg P']\!]^i)$, and hence in $([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]^i) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i)$. Either way the element is a member of $([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]^i) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i)$. But then it holds that:

(21) $([\![A]\!]^i \cap [\![P]\!]^i) \subseteq ([\![A]\!]^i \cap [\![P']\!]^i \cap [\![P]\!]) \cup ([\![A]\!]^i \cap [\![\neg P']\!]^i) \subseteq w$

By Proposition 1, w then verifies $D(A)(P)$ in i.

Now assume that w verifies $D(A)(P)$ in i. We show that w verifies $D(A)(P'P)$ in i.

Since w verifies $D(A)(P)$ in i, it holds that $([\![A]\!]^i \cap [\![P]\!]^i) \subseteq w$. For w to also verify $D(A)(P'P)$ in i, it needs to hold that $([\![A]\!]^i \cap [\![P]\!]^i \cap [\![P']\!]^i) \subseteq w$. But since $([\![A]\!]^i \cap [\![P]\!]^i \cap [\![P']\!]^i) \subseteq ([\![A]\!]^i \cap [\![P]\!]^i)$, it also holds that:

(22) $([\![A]\!]^i \cap [\![P]\!]^i \cap [\![P']\!]^i) \subseteq ([\![A]\!]^i \cap [\![P]\!]^i) \subseteq w$

Thus, w verifies $D(A)(P'P)$ in i. Since i was arbitrary, this holds for all worlds in C, and hence **Witness Set Transparency** holds.

4 Conclusion

Summing up, our predictions stand as follows:

(23) For an upwards monotone determiner:
   a. Some/All/Most/At least n/... of my students stopped smoking. ▷ Some/All/-Most/At least n/... of my students used to smoke.

(24) For downwards monotone determiner:
   a. None/Few/At most n/Less than n/... of my students stopped smoking. ▷ None/Few/At most n/Less than n/... of my students didn't use to smoke.

Importantly, universal quantifiers (including 'No/None') project universal presuppositions, while existential quantifiers project existential presuppositions. Other quantifiers project presuppositions that do not rise to full universal.

Our approach then is strongly predictive, and at least some of its predictions match the currently known empirical landscape. Of course, further empirical clarification is required to evaluate the full set of predictions, and it is to be hoped that this will happen in the future.

References

[1] Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics & Philosophy*, 4:159–219, 1981.

[2] David I Beaver. *Presupposition and assertion in dynamic semantics*, volume 29. CSLI publications Stanford, 2001.

[3] Emmanuel Chemla. Presuppositions of quantified sentences: Experimental data. *Natural language semantics*, 17(4):299–340, 2009.

[4] Irene Heim. On the projection problem for presuppositions. In M. Barlow, D. Flickinger, and N. Wiegand, editors, *Proceedings of WCCFL 2*, pages 114–125, 1983.

[5] Philippe Schlenker. Anti-dynamics: Presupposition projection without dynamic semantics. *Journal of logic, language and information*, 16(3):325–356, 2007.

[6] Philippe Schlenker. Be articulate: A pragmatic theory of presupposition projection. *Theoretical Linguistics*, pages 157–212, 2008.

[7] Robert C Stalnaker. Assertion. In Peter Cole, editor, *Syntax and Semantics 9*, pages 315–332. Brill, 1978.

[8] Sonja Tiemann. *The processing of wieder ('again') and other presupposition triggers*. PhD thesis, University of Tubingen, 2014.

Preconditions, Presupposition and Projection

Craige Roberts[1*] and Mandy Simons[2]
[1] The Ohio State University & Barnard College
[2] Carnegie Mellon University
roberts.21@osu.edu, mandysimons@cmu.edu

Abstract

We provide an account of the projection behavior of a subclass of projective contents which includes the projective contents of factives and of change of state clauses. We characterize these projective contents as ontological preconditions of events, and explain both why they tend to project, and under what circumstances they do not.

1 Introduction

This paper deals with a class of projective contents which we call *background contents*, focusing on three subclasses. Two are familiar from the presupposition literature: the veridical implications of factive clauses, and the pre-state implications of change of state (CoS) predicates. We extend our account to the less discussed implications arising from selectional restrictions of predicates.

Following prior literature (Simons et al. 2010; De Marneffe, Simons & Tonhauser 2019; Degen & Tonhauser 2022), we say that an implication *projects* when the implication is understood *as a commitment of the speaker* despite being associated with material which is in the syntactic scope of an entailment canceling operator. We describe an implication arising from utterances of a given clause as *projective* if that implication regularly (but not always) tends to project.

The motivation for the current proposal is provided by two types of observation from the prior literature. The first observation is that background contents, unlike projective implications with clearly anaphoric triggers, do not require a contextually provided antecedent. The observation that factive implications are easily accommodated goes back at least to Karttunen (1974). Spenader (2003) provides a corpus analysis which supports this claim; Tonhauser et. al (2013) provide support from fieldwork in Guarani on analogous examples. That paper concludes that background implications lack a *strong contextual felicity* constraint (-SCF): from an empirical standpoint, there is no evidence that background implications must have an antecedent in, or be entailed by, the context to be updated by their trigger.

The second kind of motivating observation pertains to the language-internal and cross-linguistic regularities concerning which implications of a given trigger project (Simons 2001; Schlenker 2021). Truth conditionally equivalent expressions across languages show the same projection behavior, as do classes of predicates within and across languages. These regularities strongly suggest that the projectivity of a given content is not an arbitrary lexical fact. Further support for

[*] The authors are listed in alphabetical order. Both authors contributed equally to developing this material and writing this paper. We acknowledge the deep influence on this work of our long collaboration with David Beaver and Judith Tonhauser; we'd also like to thank Alex Warstadt for his input and participants at PhLiP 2022 for helpful comments and questions.

this comes from evidence (Tieu et al. 2019; Schlenker 2021) of non-linguistically triggered projective implications.

Given these observations, a predictive theory of the projection of background implications requires the following:

i. An account or characterization of those implications which potentially project, allowing that projective inferences may be heterogeneous
ii. An explanation of what it is for these to project and why they do
iii. An account of patterns of projection and non-projection, crucially not one that relies on the presumption of anaphoric constraints.

Our goal is to provide an account where these three components are fully integrated with one another; that is, where the account of what projection *is* is concordant with the nature of the projective implications, and where the account of patterns of projection/non-projection is concordant with the underlying explanation of projection. Recognizing the potential heterogeneity of projective background contents, we focus here on the three classes of implications noted above, which we believe all fall under a single analysis. Specifically, we propose that:

i. The projective contents of factives and CoS preds, as well as the implication that selectional restrictions of verbs are satisfied, are *ontological preconditions* of the associated events.
ii. Projection constitutes attribution to the relevant doxastic agent, typically the speaker, of the assumption that these preconditions hold. Call this attribution global accommodation.
iii. Non-projection, including in standard filtering environments, arises when the hearer has reason *not* to take the speaker to intend global accommodation, and when an alternative interpretation is available.

2 Projective Implications and Event Preconditions

We borrow from the literature on ontological dependence (see e.g. Tahko & Lowe, 2020) the notion of an *ontological precondition*: An ontological precondition of a specific eventuality E is an object or eventuality on whose existence the existence of E is dependent. Within the broad class of ontological preconditions of an eventuality, we can distinguish preconditions with different sources. Consider for example the event of Julius Caesar dying. One precondition of this event is that Julius Caesar was previously alive: for any *a*, a precondition on the event of *a* dying is that *a* was alive. In addition, the specific event of Julius Caesar dying requires that the specific individual Julius Caesar existed; this, by at least some accounts, has as an ontological precondition that the parents of Julius Caesar existed.[†] This precondition has its source not in the fact that the event is a dying event, but in the fact that the event involved Julius Caesar. Other dying events lack this precondition, while other events involving JC (such as JC eating) share it. The projective implications of predicates are ontological preconditions of the *event type* characterized by that predicate. To say that events of type *E* are ontologically dependent on events of type *F* is to say that any event of type *E* can exist only if some appropriately situated event of type *F* exists. A speaker who evokes an eventuality *e* of type *E* is taken to be committed to the existence of instances of any eventuality types $F_1 \ldots F_n$ which are ontological preconditions of events of type *E*.

Knowledge of the ontological preconditions of eventuality types is an element of world knowledge, not linguistic knowledge. But this knowledge has linguistic consequences. If S1 describes an eventuality *e*, and S2 describes a precondition of *e*, then S1 entails S2; however, not

[†] Thanks to W. Starr (p.c.) for this observation.

all entailments of a sentence describe preconditions. We propose here two diagnostics which enable us to distinguish between entailments which characterize preconditions, and those which do not.

- *Diagnostic 1: which partly allows/allowed for*

Diagnostic frame, with target clause φ and target implication ψ: **ψ, which partly allows/allowed for φ**
Test**:** If ψ is an ontological precondition of φ, the sentence is judged true; if ψ is implied by φ but is not a precondition of it, the sentence is judged false.

Illustrations:

(1) With factive predicate *know*: target implication (the denotation of the complement of *knows*) *'*Strasbourg is in France*'*; control implication (non-projective entailment of *knows*) 'Yasmin believes that Strasbourg is in France':
   (a) Strasbourg is in France, which partly allowed for Yasmin to know that Strasbourg is in France. (T)
   (b) Yasmin believed that Strasbourg is in France, which partly allowed for Yasmin to know that Strasbourg is in France. (F)

(2) With change of state predicate *fall off*, target implication (pre-state '[agent] was on the ladder'), control implication (non-projective post-state, '[agent] was off the ladder'):
   (a) Jane was on the ladder, which partly allowed for Jane to fall off the ladder. (T)
   (b) Jane was off the ladder, which partly allowed for Jane to fall off the ladder. (F)

- *Diagnostic 2: counterfactual*

Diagnostic frame, with target predicate (with its complement) ___, target implication ψ:

a. where predicate is telic:

If not-$\psi_{subjunc}$, it would not have been possible for [agent] to ___

b. where predicate is atelic:

if not-$\psi_{subjunc}$, it would not have been possible for [agent] to come to __

Addition of *come to* ___ shifts to a telic interpretation of the predicate. Where both clauses contain atelics, the eventualities are understood as co-temporal (Partee 1984; Dowty 1986), forcing a "definitional" reading and undermining the test.

Test**:** If ψ is an ontological precondition of the target sentence, the counterfactual is judged true; if ψ is implied by φ but is not a precondition of it, the counterfactual is judged false

Illustrations:

(3) With change of state predicate *fall off the ladder*, target implication ψ_1 '[agent] was on the ladder', control implication ψ_2 '[agent] was off the ladder':
   (a) If Jane had not been on the ladder, it would not have been possible for Jane to fall off the ladder. (T)
   (b) If Jane had not been off the ladder, it would not have been possible for Jane to fall off the ladder. (F)

(4) Illustration with atelic factive predicate *knows*, complement *that Strasbourg is in France*, target implication ψ_1 'that Strasbourg is in France', control implication ψ_2 'that [agent] believed that Strasbourg is in France':

(a) if Strasbourg were not in France, it would not have been possible for Yasmin to come to know that Strasbourg is in France (T)
(b) If Yasmin did not believe that Strasbourg is in France, it would not have been possible for Yasmin to come to know that Strasbourg is in France. (F)

Space restrictions preclude a more detailed discussion of these diagnostics. What is crucial to observe here is that the diagnostics distinguish between different subtypes of entailment, and that those entailments which are diagnosed as preconditions are exactly those which are projective.

3 Projection and non-projection

We now turn to the question of projection, parts ii. and iii. of our account as laid out in the introduction. We address the following questions:

Q1: What does it mean for preconditions to project, and why do they do so?

Q2: Why do they sometimes not project?

The leading intuition underlying our proposal is articulated by Stalnaker (1974: 55): "The propositions that P and that Q may be related to each other, and to common beliefs and intentions, in such a way that it is hard to think of a reason that anyone would raise the question whether P, or care about its answer, unless he already believed that Q." We propose that one such relation is precisely the relation of ontological precondition: in the default case, where Q is an ontological precondition of P, a hearer who observes that the speaker has (typically implicitly) raised the question whether P will attribute to the speaker the assumption that Q holds, for in the default case, the question whether P is a pragmatically uninteresting one in the absence of that assumption. A more detailed discussion of the notion of a pragmatically "good" question is offered in Warstadt (2022). (We assume for current purposes a notion of question-raising drawn from Searle (1969: 124–126), according to which to predicate *P* of a subject *a* is to raise the question whether *P*(a), a notion distinct from the focus-driven sense of question-raising adopted in Roberts (1996) and much subsequent work.)

An important consequence of our view of projection in these cases is that we explain the frequently-observed generalization that accommodation is by default *global* (Gazdar 1979; Heim 1983). In our view, accommodation is a move made by an agent – here, the addressee – to support the goals of another – the speaker – the goal being to contribute appropriately to a discourse. Accommodation is not, in general, the repair of a faulty context of utterance (Thomason 1990; Roberts 2015). In the cases under consideration, the default accommodation strategy is to take the speaker to be assuming preconditions to hold. Consequently, the effect is that of global accommodation.

Nonetheless, there are cases in which it is pragmatically *implausible* that the speaker is assuming the preconditions to hold. It is precisely in such cases that the sentences in question are given non-projective readings – but again, only where those non-projective readings are themselves pragmatically plausible. There are three classes of cases where projective readings are dispreferred: (a) where the precondition is evidently taken by (some of) the interlocutors to be false or controversial, (b) where there is other evidence that the speaker is unlikely to take the precondition to be true and (c) where the precondition is at-issue. We briefly illustrate these three cases in turn.

- *Precondition understood to be false/controversial*

(5) [Context: A & B are American siblings with opposed political views. It is common knowledge between them that B believes that Trump lied about the election being stolen, and that A believes Trump. The two generally avoid discussing politics, but now they are

discussing the fact that B watched all of the public hearings of the January 6th committee, while A watched none.]
A: If Trump testifies live, I'll watch THAT hearing.
B: If the committee discovers that there is documentary evidence that Trump knowingly lied to the public, will you change your vote?

If A were to take B to presume that there is documentary evidence that Trump lied, and to intend A to so interpret the utterance, then B would be attempting a linguistic move that B knows that A would reject. An alternative interpretation is available, where the existence of the evidence is merely locally entailed (in the antecedent), and not presumed. Given A & B's history of trying not to antagonize each other, it is most plausible that B intends to be so interpreted.

Note that on this account, there is no need for "local accommodation," as there is no contextual constraint to satisfy. There is merely local entailment, due to the lexical meaning of *discover*.

- *Evidence of speaker non-commitment*

(6) A: I heard that John got drunk last night.
B: I strongly doubt that. Mary would divorce him if she discovered he was drinking.

The speaker cannot consistently simultaneously *doubt* that John got drunk and *assume* that John got drunk. So A does not take B to be making that assumption; the implication of John drinking is only held locally true. Cases characterized by Simons (2001) as *explicit ignorance cases* are similar:

(7) [Context: interlocutors are observing a stranger who is acting irritably and chewing on a pencil. They are speculating about her behavior.]
A: Maybe she just stopped smoking. That's how I acted when I quit.

Importantly, the standard filtering environments also turn out to be subcases of this effect.

(8) If Sam was in her office today, she left.
(9) Either Sam was never in her office today, or she left.

Modulo other contextual information, felicitous use of a conditional antecedent or of a disjunction signals lack of speaker commitment to the truth of the clause (cf. Gazdar 1979). In each case the non-projective reading, where the preconditions are merely locally entailed, is preferred.

- *Precondition is at-issue*

Multiple publications (Simons et al. 2010; Beaver et al. 2017; Simons et al. 2017) have argued that the projective contents of factives and change of state verbs fail to project when they are at-issue; experimental evidence for this claim is provided by (Amaral & Cummins 2015; Tonhauser, Beaver & Degen 2018). The observation is unsurprising given the view of projection we develop here: a speaker cannot reasonably assume something true when the question of its truth is currently under discussion in the discourse. For example:

(10) [Context: The interlocutors know that B is one of Jane's closest colleagues, with whom she shares details about her professional life.]
A: Is there a decision yet on Jane's tenure case?
B: I was just talking to Jane, and she isn't aware that there's been a decision.

The complement of *aware* would constitute an answer to A's question. As B has not given a direct answer, she signals that she is unable to do so. But if she is unable to do so, she also cannot

coherently presume the answer to the question; hence the hearer must attribute a non-projective reading.

In the examples above, given the non-plausibility of a projective reading, the hearer treats the relevant content as merely locally true. In other cases, where the semantics of a sentence provide an *intermediate* context to which a precondition could be accommodated, we predict that this accommodation will occur when it supports the plausibility of an interpretation. This requires more detailed discussion, but we limit ourselves here to a single example:

(11) [Context: A/B know that Yasmin's partner Phil is a US citizen]
A: Is Yasmin a US citizen?
B: I'm not sure. But if she marries Phil, then she could easily become a US citizen.

Global accommodation is ruled out here by evidence of speaker non-commitment. Accommodation into the *if*-clause provides a plausible interpretation.

4 Conclusion

The orthodox approach to presupposition and projection due to Heim (1983) and van der Sandt (1992) proposes a unified approach covering both anaphoric presuppositions and those of interest here, using lexical presupposition triggering and an algorithm for projection and filtering. All presupposition triggers are treated as anaphoric, said to impose a felicity constraint on prior context. This approach works well for the anaphoric triggers including pronouns and *too*. But there are several features of the background implications of the class of triggers of interest here which are problematic for the unified anaphoric approach, problems avoided with the preconditional account we offer above:

- The anaphoric approach predicts that all presupposition triggers always impose a strong contextual felicity constraint on prior context, hence cannot account for the fact that background content is quite generally felicitous when informative. Our account does not.
- The unified approach incorrectly predicts that embedded background content will project unless entailed by an intermediate context.
  - As a corollary, in complex sentences like conditionals, the standard projection algorithm predicts unattested presuppositions for non-anaphoric triggers, giving rise to the so-called *proviso problem* (Geurts 1996). On the present approach, there is no proviso problem because there is no projection algorithm predicting conditional presuppositions. Recent work (Silk 2022) argues that this is the correct prediction.
- The unified approach, with its semantically driven projection algorithm, cannot satisfactorily explain when and why a background implication *fails to project*. In contrast, because here projection is pragmatic, not semantic, we predict that it occurs only as long as the result is felicitous. The determination of whether content is accommodated, and if so at what level, is based on independently motivated features of the context of utterance and independently supported pragmatic constraints on discourse coherence.
- Nor can the orthodox approach explain why we would expect to find the same projective content in near-translation equivalents across languages. We expect this because these projective contents are preconditions on the respective associated event-types, hence triggered by the pragmatics of the shared meanings, not linguistic triggers *per se*.
- Because selectional restrictions are clearly non-anaphoric, their projectivity cannot be addressed by the unified approach, despite the fact that it is well-attested and strikes us as clearly parallel to that of the factives and CoS predicates.

One nice feature of the Heim/van der Sandt approach is that their projection algorithm correctly predicts parallel configurations for local anaphora resolution and for non-global accommodation of background implications. But this can be attributed to general features of dynamic local context update, as independently attested for anaphora, for Conventional Implicatures anchored under quantification (Amaral, Roberts & Smith 2007) and for intrusive implicature (Kadmon 2001). Nothing special needs to be stipulated to explain this parallel, a point we intend to elaborate on in future work.

Several other questions remain for future work. One is to address the issue of variance in projectivity demonstrated by Tonhauser et al. (2018) and Degen & Tonhauser (2022). Another is to explore how the pragmatic approach to projection can be extended to explain the attested projectivity of non-veridical clause embedding predicates such as *believe* (see Simons et al. (2017) for a preliminary account). And finally, we are exploring how to extend the precondition account beyond ontological preconditions to include, for example, preconditions on felicity, which would allow us to extend the account to the projectivity of Searlean preparatory conditions on felicity observed by Kadmon (2001).

References

Amaral, Patrícia & Chris Cummins. 2015. A Cross-Linguistic Study on Information Backgrounding and Presupposition Projection. In Florian Schwarz (ed.), *Experimental Perspectives on Presuppositions* (Studies in Theoretical Psycholinguistics), 157–172. Cham: Springer International Publishing. doi:10.1007/978-3-319-07980-6_7. https://doi.org/10.1007/978-3-319-07980-6_7 (25 November, 2022).

Amaral, Patricia, Craige Roberts & E. Allyn Smith. 2007. Review of The Logic of Conventional Implicatures by Chris Potts. *Linguistics and Philosophy* 30(6), 707–749. doi:10.1007/s10988-008-9025-2.

Beaver, David I., Craige Roberts, Mandy Simons & Judith Tonhauser. 2017. Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics* 3. Annual Reviews, 265–284.

De Marneffe, Marie-Catherine, Mandy Simons & Judith Tonhauser. 2019. The commitmentbank: Investigating projection in naturally occurring discourse. *proceedings of Sinn und Bedeutung*, vol. 23, 107–124.

Degen, Judith & Judith Tonhauser. 2022. Are there factive predicates? An empirical investigation. *Language* 98(3). Linguistic Society of America, 552–591. doi:10.1353/lan.0.0271.

Dowty, David R. 1986. The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and Philosophy* 9(1), 37–61. doi:10.1007/BF00627434.

Gazdar, Gerald. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. Academic Press.

Geurts, Bart. 1996. Local Satisfaction Guaranteed: A Presupposition Theory and Its Problems. *Linguistics and Philosophy* 19(3). Springer, 259–294.

Heim, Irene. 1983. On the Projection Problem for Presuppositions. *West Coast Conference on Formal Linguistics*, 249–260.

Kadmon, Nirit. 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*. 1st edition. Malden: Wiley-Blackwell.

Karttunen, Lauri. 1974. Presuppositions and Linguistic Context. *Theoretical Linguistics* 1, 181–194.

Partee, Barbara H. 1984. Nominal and Temporal Anaphora. *Linguistics and Philosophy* 7(3). Springer, 243–286.

Roberts, Craige. 1996. Information Structure: Towards an integrated theory of formal pragmatics. (Ed.) Yoon & Kathol. *OSU Working Papers in Linguistics* 48, 91–136.

Roberts, Craige. 2015. Accommodation in a Language Game. In Barry Loewer & Jonathan Schaffer (eds.), *A Companion to David Lewis*, 345–366. Hoboken, NJ: Wiley.

Schlenker, Philippe. 2021. Triggering Presuppositions. *Glossa: a journal of general linguistics* 6(1). Open Library of Humanities. doi:10.5334/gjgl.1352. https://www.glossa-journal.org/article/id/5415/ (16 August, 2022).

Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139173438. https://www.cambridge.org/core/books/speech-acts/D2D7B03E472C8A390ED60B86E08640E7 (29 November, 2022).

Silk, Alex. 2022. Challenges for independence-driven and context-repair responses to the proviso problem. *Natural Language Semantics* 30(3), 363–377. doi:10.1007/s11050-022-09194-6.

Simons, Mandy. 2001. On the Conversational Basis of Some Presuppositions. *Semantics and Linguistic Theory* 11(0), 431–448. doi:10.3765/salt.v11i0.3099.

Simons, Mandy, David Beaver, Craige Roberts & Judith Tonhauser. 2017. The best question: Explaining the projection behavior of factives. *Discourse processes* 54(3). Taylor & Francis, 187–206.

Simons, Mandy, Judith Tonhauser, David Beaver & Craige Roberts. 2010. What projects and why. *Semantics and linguistic theory*, vol. 20, 309–327.

Spenader, Jennifer. 2003. Factive presuppositions, accommodation and information structure. *Journal of Logic, Language and Information* 12(3). Springer, 351–368.

Stalnaker, Robert. 1974. Pragmatic Presuppositions. In Robert Stalnaker (ed.), *Context and Content*, 47–62. Oxford University Press.

Tahko, Tuomas E. & E. Jonathan Lowe. 2020. Ontological Dependence. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Fall 2020. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2020/entries/dependence-ontological/ (25 November, 2022).

Thomason, Richmond. 1990. Accommodation, Meaning, and Implicature. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in Communication*. MIT Press.

Tieu, Lyn, Philippe Schlenker & Emmanuel Chemla. 2019. Linguistic inferences without words. *Proceedings of the National Academy of Sciences* 116(20). Proceedings of the National Academy of Sciences, 9796–9801. doi:10.1073/pnas.1821018116.

Tonhauser, Judith, David I Beaver & Judith Degen. 2018. How Projective is Projective Content? Gradience in Projectivity and At-issueness. *Journal of Semantics* 35(3), 495–542. doi:10.1093/jos/ffy007.

Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89. JSTOR, 66–109.

Van der Sandt, Rob A. 1992. Presupposition Projection as Anaphora Resolution. *Journal of Semantics* 9(4), 333–377. doi:10.1093/jos/9.4.333.

Warstadt, Alex. 2022. Presuppositions, Context and Information Theory. Paper presented at the Amsterdam Colloquium.

Clause-internal coherence as presupposition resolution*

Kelsey Sasaki and Daniel Altshuler

University of Oxford, Oxford, United Kingdom
{kelsey.sasaki, daniel.altshuler} @ling-phil.ox.ac.uk

1 Introduction

(1) exemplifies what [12] coins *Clause-Internal Coherence* (CIC): inferences that arise from a single clause, despite being characteristic of (multi-clause) discourse.[1] In particular, a causal inference is salient: by sticking a knife into her husband, the described subject became a widow.

(1) A widow stuck a knife into her husband. ([3])

Recently, [9] argued that familiar pragmatic tools cannot account for CIC and proposed a previously unidentified type of enrichment, *eliciture*, that is characterized by its non-local character, familiar from research on coherence in intersentential discourse (see [16], [14] for recent overviews). One major outstanding question is how to compositionally derive clause *internal* coherence from clause *external* coherence relations.

This paper take strides in answering this question. We explore causal inferences with deverbal adjectives, comparing predicative uses in discourse like (2a)-(2b) with attributive, clause-internal uses like (2c)-(2d). §2 offers evidence from offline experiments that (i) attributive deverbals trigger causal inferences, but do so more weakly than their predicative counterparts; (ii) attributive non-deverbals behave similarly to deverbals, but are even weaker causal triggers.

(2) a. *Discourse effect-cause*: A child was **drenched**. She got hit by a big water balloon.
b. *Discourse cause-effect*: A big water balloon hit a child. She was **drenched**.
c. *Clause effect-cause*: A **drenched** child got hit by a big water balloon.
d. *Clause cause-effect*: A big water balloon hit a **drenched** child.

In §3, we explore the possibility of extending an analysis of presupposition in SDRT ([4]) to explain the experimental findings. This is motivated by the observation that CIC can arise as a presupposition, with deverbal adjectives as presupposition triggers:

(3) a. It's not the case that a drenched child got hit by a big water balloon. She was pushed into the pool.
b. A few children at Camp Hope showed up to dinner drenched. If a drenched child got hit by a big water balloon, then someone smuggled such balloons into the camp.

The analysis we propose views presupposition as a species of anaphora resolution ([22] and [18])). We show how a causal inference follows from the resolution of a coherence relation that binds the presupposed information and an attachment point that allows for projection. The salience of a causal inference, we argue, follows from the interaction of independently motivated, default axiom schemata for inferring particular coherence relations.

Finally, in §4, we summarize our contributions and questions for further research.

*Thanks to Julian Schlöder, Matt Husband, Runyi Yao, Adrian Brasoveanu, Bridget Copley, Fabienne Martin, Roger Schwarzschild, audiences/reviewers at SuB27, XPRAG9, and participants at COCOA. Any errors are ours.

[1] [12], [13] and [9] also use 'CIC' to characterize multi-clause examples like *The company fired the manager who was embezzling money.* Here, we use 'CIC' for examples that (at least prima facie) involve a single clause.

2 Experimental support for clause-internal coherence

2.1 Experiment 1: Deverbal adjectives

We hypothesized that attributive deverbal adjectives can trigger clause-internal inferences that parallel those that predicative deverbals trigger at the discourse level. To test this, and to probe for effects of (linear) cause/effect order, we used a 2x2 design crossing INFERENCE DOMAIN {DISCOURSE, CLAUSE} with CAUSE/EFFECT ORDER {CAUSE-EFFECT, EFFECT-CAUSE} for 40 items, as in (2). 42 filler items were balanced for causal inference strength (strong/medium/weak). In an initial Likert-scale study, we found evidence that attributive deverbals can trigger causal coherence-type inferences, but more weakly than their predicative counterparts.[2]

Building on our initial findings, we hypothesized that, in clause-internal contexts, at least two coherence inferences—a causal inference (4a) and a non-causal inference (4b)—compete for salience, and that the less-salient inference is nonetheless available. In discourse-level contexts, on the other hand, we did not expect the causal inference to have any significant competition, either in salience or general availability. We tested this via a two-stage forced-choice task, under the linking hypothesis that the first stage gauged the relative saliencies of the provided interpretations, and that the two stages combined gauged the overall availability of each interpretation. Participants[3] (n=48) first chose between a causal and a non-causal interpretation, e.g., the options in (4) for the item in (2). On the same screen, participants then answered the question, *Is the other option also a reasonable description of what happened?*

(4) a. The child was drenched because she got hit by the big water balloon.
b. The child was already drenched when she got hit by the big water balloon.

	Causal			Non-causal		
	1st Choice	2nd Ch.	Overall Avail.	1st Ch.	2nd Ch.	Overall
Disc. Eff.-Cause	0.91	0.07	0.98	0.09	0.33	0.42
Disc. C-E	0.95	0.03	0.98	0.05	0.19	0.24
Clause. E-C	0.61	0.25	0.86	0.39	0.32	0.71
Cl. C-E	0.61	0.24	0.85	0.39	0.33	0.72

Table 1: Rates of interpretative choices for deverbal adjectives

The proportions of causal and non-causal choices are summarized in Table 1. Data were analyzed in R with maximal Bayesian mixed effects linear regression models [7, 8]. The causal interpretation was a both more frequent first choice (3.89, [2.78, 5.16]) and more available overall (2.65, [1.21, 4.33]) in DISCOURSE conditions, compared to CLAUSE conditions. With respect to the overall availability of the non-causal interpretation, we found significant main effects of both factors: overall availability was higher for CLAUSE conditions than DISCOURSE conditions (-2.54, [-3.16, -1.97]) and higher for EFFECT-CAUSE order than CAUSE-EFFECT (-0.53, [-0.82, -0.24]). There was also a significant interaction: in DISCOURSE conditions, the non-causal interpretation was more available overall in the EFFECT-CAUSE condition than the CAUSE-EFFECT condition (-1.18, [-1.56, -0.82]), but no difference obtained between the CLAUSE conditions.

These results suggest that: (i) attributive deverbal adjectives can reliably trigger inferences that parallel discourse-level coherence inferences, but (ii) attributive deverbals are less likely to

[2]For full details of this study and a single-stage forced choice study, see [19].
[3]Participants were native-English-speaking, UK-based Prolific workers. Experiments ran on PCIbex [23].

trigger causal coherence inferences than their predicative counterparts. This may be because a non-causal inference can more readily compete with the causal inference in clause-internal contexts, though in discourse this inference is at least available in many of the tested cases.

2.2 Experiment 2: Non-deverbal adjectives

Expt. 1 focused on deverbal adjectives in order to elicit the most robust clause-internal coherence inferences; deverbal adjectives, while not overtly clausal, are derivationally related to verbs and describe events. In Expt. 2 (n = 60), we tested whether adjectives lacking these properties also trigger causal inferences when used attributively. The task was the same as in Expt. 1, but the critical adjectives were non-deverbal—e.g., *wet* instead of *drenched* in (2). The results are summarized in Table 2. Data were analyzed in the same way as the Expt. 1 data.

	Causal			Non-causal		
	1st Choice	2nd Ch.	Overall Avail.	1st Ch.	2nd Ch.	Overall
Disc. Eff.-Cause	0.86	0.04	0.90	0.14	0.31	0.45
Disc. C-E	0.90	0.03	0.93	0.10	0.32	0.42
Clause. E-C	0.46	0.18	0.64	0.54	0.16	0.70
Cl. C-E	0.44	0.20	0.64	0.57	0.13	0.70

Table 2: Rates of interpretative choices for non-deverbal adjectives

We found that the causal interpretation was a more frequent first choice for DISCOURSE conditions than CLAUSE conditions [3.29, (2.70 3.92)]. There was also a significant interaction: between the DISCOURSE conditions, the causal first choice was more frequent for the CAUSE-EFFECT order (0.65, [0.11, 1.26]), but the CLAUSE conditions did not differ from one another. The causal interpretation was also more available overall for the DISCOURSE conditions (2.55, [1.99, 3.17]). The non-causal interpretation was both a more frequent first choice and more available overall (-1.40, [-1.76, -1.04]) in the CLAUSE conditions than the DISCOURSE conditions.

We also analyzed the combined results of Expts. 1-2, with ADJECTIVE TYPE as an additional factor. We found an effect of ADJECTIVE TYPE such that the causal interpretation was the more frequent first choice (1.16, [0.44, 1.94]) and more available overall (2.03, [1.23, 2.91]) with deverbals; the non-causal interpretation was the more frequent first choice for non-deverbals, and was in fact the first choice over half the time for non-deverbals, but not for deverbals. However, ADJECTIVE TYPE did not have a significant main effect on its overall availability.

The results of Expt. 2 suggest that non-deverbal adjectives can trigger coherence inferences, even when attributive. However, regardless of context, the causal inferences triggered by non-deverbals seem to be both less salient and less available overall than those triggered by deverbals.

3 Formal Proposal

3.1 Analyzing deverbal adjectives

Discourse coherence. We first consider (5), which illustrates how the discourse-level inference in (2a) is established. According to the asserted content in π_a, there is a child x who is in a drenched state s, which, we propose, was caused by some event e. According to the asserted content in π_b, there is a water balloon y, which was used in a hitting event e', whose theme is

some individual z. By establishing the coherence relation Elaboration[4] between π_a and π_b, e is resolved to e' and z is resolved to x.[5] This resolution entails a causal link between π_a and π_b, although Elaboration is a non-causal relation. The relevant paraphrase is, 'A drenched state that some child was in was caused by some event; that event is a water-balloon hitting event'.

(5)

π_a		π_b	
π_a:	x, s, e $\texttt{child}(x)$ $\texttt{drenched}(s)$ $\texttt{in}(s, x)$ $\texttt{cause}(e, s)$ e = ?	π_b:	y, z, e' $\texttt{water.balloon}(y)$ $\texttt{hit}(e')$ $\texttt{with}(e', y)$ $\texttt{theme}(e', z)$ z = ?

Let's now turn to the non-causal interpretation of (2a): 'A water balloon hit an already drenched child; some other event brought about the drenched state.' This interpretation follows from establishing Background, which entails that the eventualities described by its arguments overlap in time. For (5), Background ensures that s overlaps e' and that e is bound, but, crucially, not resolved to e'. Assuming that a coherence relation that leaves a discourse referent (dref) unresolved is dispreferred to one that resolves all drefs, our formalization is consistent with the Expt. 1 finding that the causal interpretation is preferred for (2a).[6]

As for (2b), we again have the representational content in (5), but π_a and π_b are interpreted in reverse order. This means that the event causing the drenched state is resolved anaphorically, instead of cataphorically, i.e., the second sentence of (2b) doesn't expand on the balloon-hitting event, but rather describes its result state. Elaboration is ruled out; we propose that the operative coherence relation here is Continuation[7]. It is non-causal like Elaboration, but supports an anaphoric resolution in which the event causing the drenched state is identified with the balloon-hitting event, thus deriving the causal inference. Finally, while a non-causal inference is not prominent in (2b), it can nevertheless follow from establishing Background, as in (2a).

Clause-internal coherence. We now show how an SDRT analysis of presupposition [4] can be extended to capture (2c) given our hypothesis that attributive deverbals are presupposition triggers. Consider (6). According to the asserted content in π_a, there is an event e of being hit by a water balloon and a child x is e's theme. According to the presupposed content in π_b (contributed by the deverbal), there is a drenched state s, which was caused by an event e', and which holds of an individual z. The presupposed content is related to an underspecified attachment point u by an underspecified coherence relation R. Following our analysis of (2a), we can resolve R to Elaboration or Background, with u being resolved to π_a in either case.[8] As before, Elaboration leads to a causal interpretation, and Background to a non-causal one.

As for (2d), its representation is the same as (6), with one important distinction: the first argument of R is now u (which is resolved to π_a). Following our analysis of (2b), we can

[4]Establishing Explanation instead would result in a bizarre interpretation: the water-balloon hitting event is the reason for a drenched state being caused by the water-balloon hitting event. However, Explanation is crucial for adjectives that don't semantically contribute $\texttt{cause}(e, s)$, e.g., non-deverbal adjectives.

[5]Here we assume that establishing rhetorical relations and resolving the interpretation of a context sensitive expression are correlated tasks (see, e.g., [11, 17, 15, 21]). The semantics of Elaboration ensures that its second argument provides more information about the *same event* described by its first argument (see [5]).

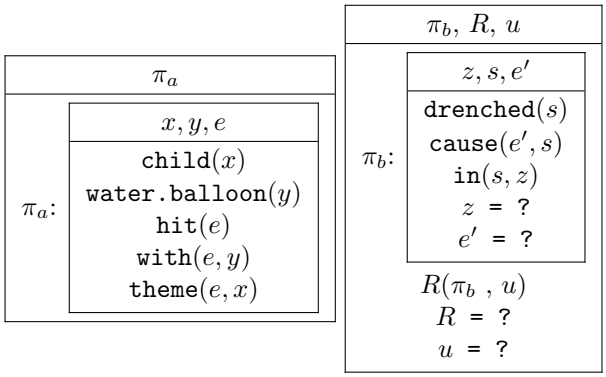
[6]At least one other factor likely underpins this finding: an interpretative default, well-established in experimental psychology [10, 24], to infer a causal link between adjacent eventualities whenever possible.

[7]Establishing Result here would lead to the same bizarre interpretation that was discussed in fn.4.

[8]π_a is thus either a cataphoric presupposition or a postsupposition (see [6]). We are agnostic about which.

resolve R to either Continuation or Background. As before, Continuation leads to a causal interpretation, while Background leads to a non-causal one.

(6)



We are now ready to address two core observations from Expt. 1: (i) causal inferences are more salient and more available in (2a)-(2b) than (2c)-(2d); (ii) for each condition, causal inferences are more salient and available than non-causal inferences. We propose that (i)-(ii) can be captured by assuming two axioms that help adjudicate between possible relations[9]:

(7) *Resolve drefs*: Establish the relation that produces the least unresolved drefs.

(8) *Constraint on presuppositions*: If possible, resolve R with Background.

While the discourse-level Elaboration and Continuation respect both axioms, their clause-internal counterparts both violate (8). Yet, despite violating (8), the presuppositional Elaboration and Continuation were still the preferred interpretations in Expt. 1. We can derive this by further assuming that (7) takes precedence over (8).

Our proposal does not presently capture the experimental finding that non-causal inferences are more available overall for (2c) and (2d) than for (2a) and (2b).

3.2 Analyzing non-deverbal adjectives

Let us finally consider non-deverbal adjectives like *happy*. First, we adopt an independently motivated axiom from Schlöder [20, Ch.7]:

(9) *Schlöder's causal axiom*: Given a pair of eventuality descriptions α, β:
   a. if it's possible that the eventuality described by α caused the eventuality described by β, establish Result(α, β).
   b. if it's possible that the eventuality described by α was caused by the eventuality described by β, establish Explanation(α, β).

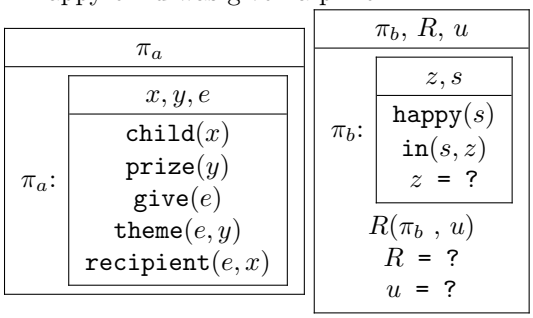
We assume that non-deverbals differ from deverbals in not semantically contributing cause(e, s), but that, like deverbals, they presuppose the described state. The latter is evidenced by (10), which shows the expected projection behavior. Thus, we analyze (11a) as in (11b).

(10) a. It's not the case that a happy child was given a prize. She was given a new toy.
   b. A few children at Camp Hope showed up to dinner happy. If a happy child was given a prize, then someone smuggled the prize into the camp.

[9] While (7) is straightforward, we note that (8) is motivated by the intuition that presuppositions are not-at-issue or backgrounded content (see, e.g., [1] and references therein).

(11) a. A happy child was given a prize.

b.



According to the asserted content in π_a, there is a child x who is the recipient of a giving event e, whose theme is a prize y. According to the presupposed content in π_b, there is a state s of being happy that holds of an individual z. This content is related to an attachment point u by a coherence relation R. Unlike in (2c), Elaboration is not a possibility since non-deverbal adjectives don't, by assumption, presuppose cause(e, s). However, precisely because they don't presuppose cause(e, s), Explanation is possible (recall fn.4), which derives the causal interpretation: the child's happiness is explained by the prize that they were given. This respects Schlöder's axiom, but violates (8). Of course, we can also derive a non-causal interpretation via Background, which would violate Schlöder's axiom but respect (8).[10] Recall that Expt. 2 revealed that the non-causal interpretation (with Background) is the more salient and more available reading for (11a), suggesting that (8) outweighs Schlöder's axiom.[11]

Let us now consider (12). This example would analyzed as in (11b), except that the first argument of R would now be u (resolved to π_a). We can derive the non-causal reading by resolving R to Background, and the causal reading by resolving R to Result. Our findings for (12) suggest, as they did for (11a), that (8) outweighs Schlöder's axiom.

(12) A prize was given to a happy child.

Finally, we note that in the discourse versions of (11a)/(12), the axiom in (8) isn't operative. Given Schlöder's axiom, a causal inference is thus most salient via Explanation/Result.

4 Conclusion

We have provided experimental support for the existence of CIC via offline evidence that attributive (non-)deverbal adjectives can trigger the same causal inferences within clauses that their predicative counterparts can trigger across clauses, albeit more weakly. Our forced choice methodology probes whether speakers accept interpretations that are spelled out for them, not which interpretation(s) they might independently infer. Future experiments will investigate what kinds of coherence inferences, if any, speakers draw in the absence of explicit prodding.

To explain the experimental results, we used tools in SDRT, which allowed us to show that causal inferences can be derived in various ways, depending on whether deverbal adjectives are used attributively or predicatively. If the former, they are presupposition triggers and Elaboration/Continuation competes with Background; if the latter, Explanation/Result competes with Background. These different competitions (cashed out in terms of interaction between default axioms) correlate with the difference in the relative salience of the causal inferences.

[10] The axiom in (7) is not in play here, as non-deverbals do not introduce an event dref to resolve.
[11] For further discussion of Schlöder's axiom, including other constraints that it competes with, see [2].

References

[1] Márta Abrusán. Presuppositions. In Daniel Altshuler, editor, *Linguistics meets Philosophy*, pages 439–469. Cambridge University Press, 2022.

[2] Daniel Altshuler. A puzzle about narrative progression and causal reasoning. In Andreas Stokke and Emar Maier, editors, *The Language of Fiction*. Oxford University Press, Oxford, 2021.

[3] G. E. M. Anscombe. Under a description. *Noûs*, 13:219–233, 1979.

[4] Nicholas Asher and Alex Lascarides. The semantics and pragmatics of presupposition. *Journal of Semantics*, 15(2):239–299, 1998.

[5] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, Cambridge, England, 2003.

[6] Adrian Brasoveanu and Anna Szabolcsi. Presuppositional too, postsuppositional too. 2013.

[7] Paul-Christian Bürkner. brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, 80:1–28, 2017.

[8] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[9] Jonathan Cohen and Andrew Kehler. Conversational eliciture. *Philosophers' Imprint*, 21:1–26, 2021.

[10] Arthur C Graesser, Murray Singer, and Tom Trabasso. Constructing inferences during narrative text comprehension. *Psychological review*, 101(3):371, 1994.

[11] Jerry Hobbs. Coherence and coreference. *Cognitive Science*, 3:67 – 90, 1979.

[12] Jerry Hobbs. Clause-internal coherence. In Peter Kühnlein, Anton Benz, and Candace L. Sidner, editors, *Constraints in Discourse 2*, page 15–34. John Benjamins, 2010.

[13] Jet Hoek, Hannah Rohde, Jacqueline Evers-Vermeul, and Ted J.M. Sanders. Expectations from relative clauses: Real-time coherence updates in discourse processing. *Cognition*, 210, 2021.

[14] Katja Jasinskaja and Elena Karagjosova. Rhetorical relations. In Daniel Gutzmann, Lisa Matthewson, Cécile Meier, Hotze Rullmann, and Thomas Ede Zimmermann, editors, *Companion to Semantics*. Wiley, 2020.

[15] Elsi Kaiser and Boutaina Cherqaoui. Effects of coherence on anaphor resolution, and vice versa: Evidence from French personal pronouns and anaphoric demonstratives. In A. Holler and K. Suckow, editors, *Empirical Perspectives on Anaphora Resolution*, pages 51–78. 2016.

[16] Andrew Kehler. Coherence relations. In Robert Truswell, editor, *The Oxford Handbook of Event Structure*, pages 583–604. Oxford University Press, Oxford, 2019.

[17] Andrew Kehler, Laura Kertz, Hannah Rohde, and Jeffrey L. Elman. Coherence and coreference revisited. *Journal of Semantics*, 25(1):1–44, 2008.

[18] E.J. Krahmer. *Discourse and presupposition*. PhD thesis, 1995. Pagination: vii, 188.

[19] Kelsey Sasaki and Daniel Altshuler. Clause-internal coherence: A look at deverbal adjectives. In *Proceedings of Sinn und Bedeutung 27*. Charles University, Prague, to appear.

[20] Julian J Schlöder. Assertion and Rejection,, 2018. ILLC Dissertation Series, DS-2018-07.

[21] Una Stojnić. *Context-sensitivity in a Coherent Discourse*. PhD thesis, Rutgers University, 2016.

[22] Rob Van der Sandt. Presupposition Projection as Anaphora Resolution. *Journal of Semantics*, 9(4):333–377, 11 1992.

[23] Jeremy Zehr and Florian Schwarz. Penncontroller for internet based experiments (ibex). *DOI: https://doi. org/10.17605/OSF. IO/MD832*, 2018.

[24] Rolf A Zwaan, Joseph P Magliano, and Arthur C Graesser. Dimensions of situation model construction in narrative comprehension. *Journal of experimental psychology: Learning, memory, and cognition*, 21(2):386, 1995.

Presupposition Triggering Reflects Pragmatic Reasoning About Utility*

Alex Warstadt

ETH Zürich
Zürich, Switzerland
awarstadt@ethz.ch

Abstract

I present an RSA account of presupposition triggers building off prior work by Qing, Goodman, and Lassiter (2016). Intuitively, presupposition accommodation can occur through a listener's reasoning about the speaker's model of the common ground. Listeners assume that speakers try produce utterances with high contextual utility, and they will modify their own understanding of the context to rationalize low utility utterances. The account derives phenomena including accommodation of genus-species presuppositions due to Abusch (2002), QUD sensitivity of presuppositions, and gradience in the strength of some presuppositions.

1 Introduction

Theories of presupposition *triggering* (Abusch 2002; Simons, Tonhauser, et al. 2010; Abrusán 2011; Schlenker 2021) must contend with defeasibility and context sensitivity of projective inferences arising from the diverse class of *soft triggers* (Abusch 2002; Abusch 2010). This work presents a pragmatic account of triggering, building off notions of utterance utility from the Rational Speech Acts (RSA) framework (Frank and Goodman 2012; Qing, Goodman, and Lassiter 2016): I propose that a listener is more likely to accommodate a proposition p in circumstances where the speaker's utterance has higher utility if p is in the common ground. More specifically, listeners and speakers necessarily have separate models of the common ground (Stalnaker 1978), which they try to coordinate as much as possible. However, if their models of common ground diverge, sometimes the speaker may produce an utterance which they incorrectly judge will be informative to the listener. In such cases, a pragmatic listener can reason about what the speaker's model of common ground might have looked like and update their beliefs accordingly.

This basic story has been suggested previously in various forms (Stalnaker 1974; Simons 2001; Simons, Beaver, et al. 2017). Most relatedly, Qing, Goodman, and Lassiter (2016) propose an extension to the RSA model in which the pragmatic listener tries to infer not just the true state of the world, but also the speaker's model of the common ground. However, this earlier work fails to recognize some of the more interesting predictions of this kind of model. Exploring these predictions is the focus of this paper.

The Genus-Species Presupposition For the sake of illustration, I focus my discussion on genus-species presuppositions such as (1-a), first discussed by Abusch (2002). In these examples, a "species" predicate *have a green card* triggers a defeasible inference that a "genus" predicate

*I am grateful to Craige Roberts and Mandy Simons for extensive discussions on this topic. The analysis in this paper also benefited from input from Philippe Schlenker, Will Merrill, and Ethan Wilcox.

is a non-US citizen applies to Tom.[1]. As expected with presuppositions, this inference seems to project (defeasibly) out of negation (1-a) given that the Question Under Discussion (QUD) is as in (2-a). It also projects universally out of quantifiers (1-b). However, Abusch 2002) also considers another context where the local bar is giving free drinks to anyone with a green card, and the QUD is whether Tom qualifies for a free drink (2-b). In this context, (1-a) is reported to lack the genus presupposition.

(1) a. Tom doesn't have a green card.
b. None of the new hires has a green card.

(2) a. Does Tom need a visa?
b. Can Tom get a free drink?

The lexical semantic account of presupposition triggering assumed by many (e.g., Heim 1983; Schlenker 2009) struggles with soft triggers for two reasons: (a) While local accommodation provides a mechanism for blocking projection optionally, it does not explain why projection is only blocked in some special contexts; and (b) whereas the plausibility of lexical triggering is debatable for triggers like factives and change-of-state predicates, genus presuppositions are certainly not lexically encoded on species predicates (Abusch 2002; Abusch 2010; Schlenker 2021). Pragmatic approaches to these and other soft triggers have suggested that projection occurs when doing so increases the utility of the response (Simons 2001; Simons, Beaver, et al. 2017). However, the notion of utility in these works is left vague.

2 The RSA Approach to Accommodation

I adopt an earlier approach to presuppositions in RSA by Qing, Goodman, and Lassiter (2016) with a few modifications, as shown in (3).

(3) a. $L_0(Q(w)|u,C) \propto \sum_{w \in Q(w) \cap [[u]] \cap C} P(w)$
b. $S_1(u|Q,w,C) \propto exp(\alpha * Util(u||Q,w,C))$
c. $Util(u|Q,w,C) = \begin{cases} log(L_0(Q(w)|u,C)) - Cost(u) & if [[u]](w) = 1 \\ -\infty & if [[u]](w) = 0 \end{cases}$
d. $L_1(w,C|u,Q) \propto S_1(u|Q,w,C) * P(Q|w,C) * P(w|C) * P(C)$

Definitions The agents reason about a fixed set of worlds $W = \{w_1, ..., w_n\}$ and utterances $U = \{u_1, ..., u_m\}$, provided in Table 1 in the Appendix. Additionally, the model includes two contextual components: a context C and a QUD Q. Following Stalnaker's definition of the Context Set (Stalnaker 1978), a context in this model is the set of possible worlds consistent with publicly shared beliefs of the conversational agents. The QUD serves to delimits the topic of conversation (Van Kuppevelt 1995; Ginzburg 1996; Roberts 1996/2012). A QUD in this model is a partition over the set of possible worlds, where each cell represents a possible answer to the question (Groenendijk and Stokhof 1984). I write $Q(w)$ to denote the cell of the partition to which w belongs. The list of QUDs in the model is also provided in table 1.

Agents The literal listener in the model (3-a) is provided with a QUD, an utterance, and a context. They infer a probability distribution over the set of possible answers to the QUD by considering the prior probabilities of worlds consistent with the utterance and the context.

[1] A green card grants permanent residence to a non-US citizen. A US citizen cannot be a green card holder, and a green card holder does not need a visa to live or work in the US.

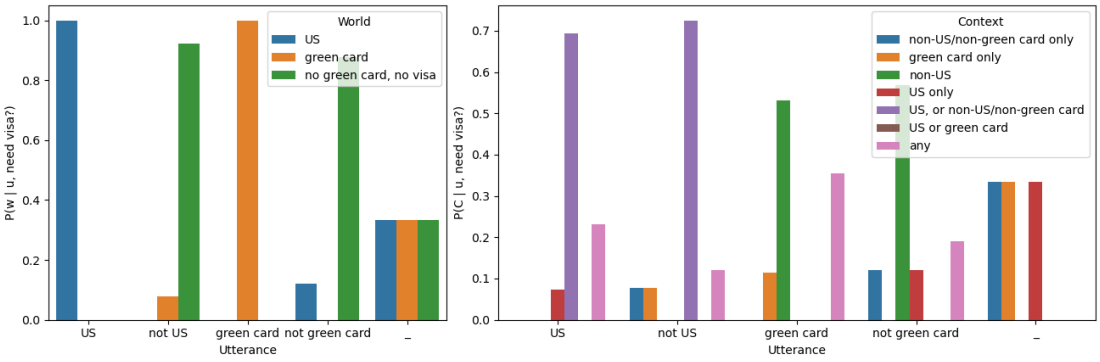


Figure 1: Pragmatic listener's posteriors over worlds (left) and contexts (right) given the QUD *Does Tom need a visa?*.

The speaker in the model (3-b) produces an utterance u with probability in proportion to the exponential of u's utility, times a rationality parameter α. The utility function is of key importance to the model. In prototypical RSA models (Frank and Goodman 2012), the speaker's goal is to reduce the listener's surprisal upon learning the actual state of the world. In this model, however, the speaker does not care to reduce the surprisal of the actual state of the world *per se*, but rather the correct answer to QUD. Furthermore, the speaker considers the reduction of surprisal with respect to C. These two factors will affect a speaker's choices in several ways. For instance, two utterances may have the same utility even if they differ in informativity overall, as long as they have the same informativity with respect to the QUD. Additionally, an utterance that has low utility in the weakest context $C = W$ may have high utility in a context C' that excludes some worlds.[2]

The pragmatic listener L_1 (3-d) differs from the prototypical RSA model in that they do joint inference over a context C in addition to reasoning over the set of possible worlds W. This is the main insight of Qing, Goodman, and Lassiter's model that I make use of. L_1 reasons about this distribution using Bayes' Rule. By assumption, $P(u|Q, w, C)$ is modeled as $S_1(u|Q, w, C)$, the speaker's probability distribution over utterances. The listener also considers three priors. First, there is a prior over QUDs $P(Q|w, C)$, which I assume to be uniform, for simplicity. Second, there is a prior over worlds given a context $P(w|C)$, which I take to be proportional to the prior over worlds $P(w)$ when $w \in C$, and 0 otherwise. Third, there is a prior over contexts C. The choice of this prior has significant implications for the predictions of the model, as I discuss in greater detail in Section 4. In the simplest case, this distribution is uniform over the set of all contexts, i.e. $\mathcal{P}(W)$.

3 Predictions

Genus-Species Presuppositions This model generates the basic genus-species presupposition. Recall that when the QUD is *Does Tom need a visa?* (2-a), the speaker's utterance

[2]The utility function also includes a condition that it returns $-\infty$ when the utterance is strictly false in the actual world. Without such an "honesty correction", the speaker will assign high utility to false utterances that happen to get the listener to believe the correct answer to the QUD. This is an interesting result which may be realistic for some speakers in some settings, and it has been leveraged to explain speakers' use of non-literal language (Kao et al. 2014). However, it is not crucial to account for projective inferences, and I find that adding this correction leads to more intuitive results.

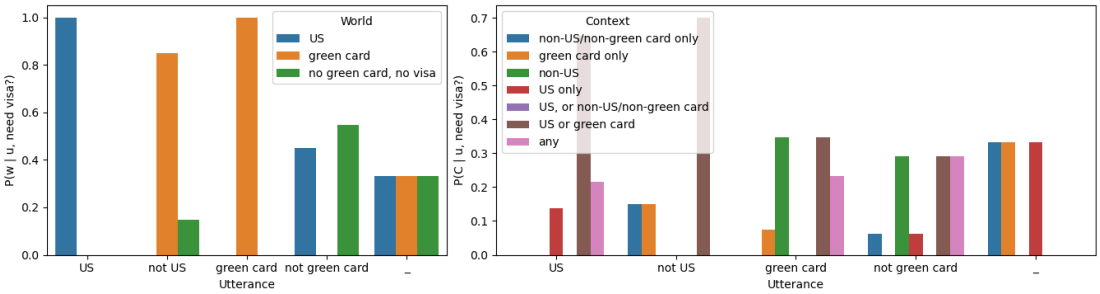


Figure 2: Pragmatic listener's posteriors over worlds (left) and contexts (right) given the QUD *Can Tom get a free drink on green card night?*.

Tom doesn't have a green card triggers an inference that Tom is not a US citizen (and therefore does need a visa). This inference is in essence a kind of relevance implicature, but the way in which the implicature is computed is not simply by reasoning about alternative utterances that the speaker could have produced, but also about alternative common grounds that the speaker could have been assuming.

The results for this model are shown in Figure 1.[3] The most interesting case is the one where the speaker's utterance is *not green card* (shorthand for (1-a)). In the listener's posterior over worlds, we can see that they assign significantly higher probability to the world in which Tom is a non-US citizen without a green card over the world in which Tom is a US citizen. This is notable, as without this kind of pragmatic reasoning, the literal listener assigns equal probability to these worlds given the same utterance.

The listener's posterior over contexts shows a related picture. When the utterance is *not green card*, the listener assigns over 50% probability mass to the context that entails that Tom is a non-US citizen. This is exactly the predicted accommodation. Interestingly, the listener still assigns a good deal of probability mass to other contexts, which may be a reasonable prediction, given the defeasibility of the genus-species presupposition.

QUD Sensitivity The model also predicts the QUD sensitivity of this inference. In particular, when the QUD is *Can Tom get a free drink on green card night?* and the utterance is *Tom doesn't have a green card*, there is little temptation to accommodate that Tom is a non-US citizen. Intuitively, this is because the speaker's utterance is an exhaustive answer in all contexts, meaning it is equally informative with respect to the QUD regardless of the common ground. In other words, there is no reasoning about the speaker's model of the common ground that will allow the listener to ascribe greater utility to the speaker's utterance.

This intuitive story is borne out in the results in Figure 2. In the listener's posterior over worlds given the utterance *no green card*, they assign nearly equal probability to the world in which Tom is a US citizen and the world in which he is a non-US citizen without a green card. Similarly, in the posterior over contexts, we see no special preference for the context that entails that Tom is a non-US citizen.

Family-Genus-Species Presuppositions I observe that the strength of Abusch's genus-species presupposition is related to the relative probability of the genus predicate and the

[3]In these and subsequent simulations, the value of α is set to 100. I found that a relatively high rationality parameter was important for generating the desired inferences.

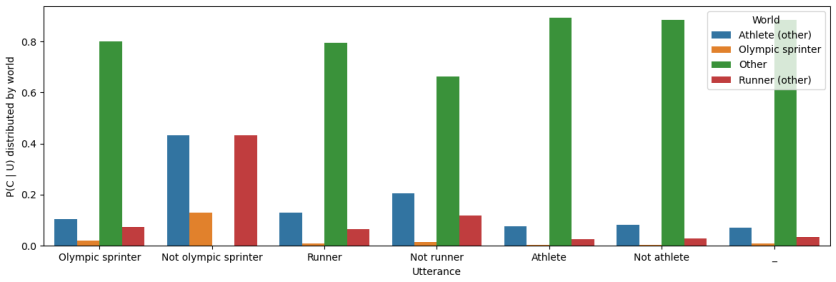


Figure 3: Pragmatic listener's posterior over contexts. Rather than plotting the probability for each of the 15 possible contexts, I plot the probability of each world according to the listener's posterior over contexts. This is not the same as the listener's posterior over worlds. Instead, given a context, I assume the probability of worlds is proportional to the world's prior probability. Thus the total probability of a world according to the listener's context posterior is the sum over all contexts of the world's probabilities in the context.

species predicate. Equivalently, rather than a pair of predicates, we can consider a triple of predicates, which I refer to as family, genus, and species. For example, if the species predicate is *Olympic sprinter*, a reasonable genus predicate is *runner*, and a reasonable family predicate is *athlete*. My observation is the projective inference that the family predicate is true (4-c) from negating the species predicate (4-a) is subjectively stronger than the projective inference from negating the genus predicate (4-b).

(4) a. Tom is not an Olympic sprinter.
b. Tom is not a runner.
c. Tom is an athlete.

This intuition fits well with the utility based story advocated above. In particular, Olympic sprinters are so rare, that there is almost no utility in asserting that Tom is not an Olympic sprinter in a typical context. Therefore, if a speaker utters (4-a), the listener may infer that the speaker is assuming a common ground in which it is significantly more likely that Tom could be an Olympic sprinter, for instance a context in which Tom is an athlete of some kind. By contrast, runners are common enough that it is typically informative to assert that Tom is not a runner. True, this assertion has even higher utility given the assumption that Tom is an athlete, but this inference is not so necessary to rationalize the speaker's utterance.

To test the model against this intuition, I use a different set of worlds and utterances from the previous examples, as illustrated in Table 2 in the Appendix. I also set a prior for each of the four worlds which reflects their approximate relative values. Note that while I do not vary the QUD in this example, the QUD can be implicitly understood to be *What is Tom's hobby*, or something similar.

Figure 3 shows the listener's probability distribution over contexts aggregated by world, with $\alpha = 1000$. For nearly all utterances, the listener infers that the speaker's model of the common ground makes it highly likely that Tom is a non-athlete. This reflects the high prior probability of this world. Crucially, this is not the case when the utterance is *Tom is not an Olympic sprinter*. In this case, the probability of the non-athlete world shifts to the worlds where Tom is a runner or another kind of athlete. Also as predicted, we see a much weaker version of this effect for the genus-family case, when the utterance is *Tom is not a runner*.

4 Discussion and Outstanding Issues

Extension to Other Triggers The success of this model for one kind of presupposition trigger (genus-species presuppositions) raises the question, "Can a similar story be told for all triggers?" For hard triggers such as additive particles like *too*, this is almost certainly not the case, as these triggers show almost no gradience or contextual sensitivity. However, there is a wide variety of other soft triggers for which this analysis is appealing, including change-of-state triggers and factives. Indeed, if we consider entities who have stopped smoking to be the "species", then entities who have smoked at some point may be considered the "genus". A classic example due to Geurts (1994) shows the contextual sensitivity of *stop* (5): Given an adequate QUD, there is less need to accommodate that Tom was ever a smoker.[4]

(5) Why is Tom chewing on his pencil? Did he recently stop smoking?

However, I am cautious to jump to the conclusion that this kind of on-the-fly pragmatic reasoning is solely responsible for the presuppositions of change-of-state predicates and factives. Intuitively, the inferences from these triggers are more indefeasible—and more conventionalized—than genus-species presuppositions. In fact, change-of-state predicates were the focus of Qing, Goodman, and Lassiter (2016)'s analysis, and they make some questionable predictions. Namely, they predict that given the QUD *Did Tom smoke in the past?*, the assertion *Tom didn't stop smoking* should give rise to a projective inference that Tom does not currently smoke, and that this inference should be just as likely as the typical initial state presupposition given the QUD *Does Tom smoke now?*.

Still, there is one parameter in the model that could potentially derive this asymmetry: the prior over contexts. If we assume that facts about Tom's past are more likely to be common ground than facts about Tom's present, this might help to explain why we more easily accommodate facts about the initial state of change-of-state predicates. A similar idea is proposed by Schlenker (2021).

Choosing What to Accommodate The analysis gives the listener great freedom to accommodate anything that might increase the utility of the speaker's utterance. In reality, listener's accommodations may be much more constrained. For instance, given the utterance *Tom is not an Olympic sprinter*, a listener could in theory accommodate that Tom has an Olympic gold medal, or that Tom has an appointment with a world famous orthopedist. Either accommodation would greatly increase the utility of the utterance, but they seem far less likely than the accommodation that Tom is a runner.

Part of the explanation must come down to the listener's prior over worlds. It is far more likely that Tom is a runner than that he is an Olympic gold medalist. If a listener is trying to find a *likely* explanation for a low utility utterance, they should accommodate facts which are more likely *a priori*. Indeed, Degen and Tonhauser (2022) do find that the likelihood of a projective inference is related to its prior probability in this way.

Future Work Future work should extend this analysis to projection of presupposition out of environments other than negation, including quantifiers, conditionals, and questions. It will also be important to test quantitative predictions of the RSA analysis against human judgments.

[4]The account could be extended to projection from interrogatives by operationalizing the utility of a question as its entropy, following van Rooy (2003) and van Rooy and Schulz (2004).

References

Abrusán, Márta (2011). "Predicting the presuppositions of soft triggers". In: *Linguistics and philosophy* 34.6, pp. 491–535.

Abusch, Dorit (2002). "Lexical alternatives as a source of pragmatic presuppositions". In: *Semantics and linguistic theory*. Vol. 12, pp. 1–19.

— (2010). "Presupposition triggering from alternatives". In: *Journal of Semantics* 27.1, pp. 37–80.

Degen, Judith and Judith Tonhauser (2022). "Prior beliefs modulate projection". In: *Open Mind* 5, pp. 59–70.

Frank, Michael C and Noah D Goodman (2012). "Predicting pragmatic reasoning in language games". In: *Science* 336.6084, pp. 998–998.

Geurts, Bart (1994). "Presupposing". PhD thesis. Universität Osnabrück.

Ginzburg, Jonathan (1996). "Dynamics and the semantics of dialogue". In: *Seligman, Jerry, & Westerst ahl, Dag (eds), Logic, language and computation* 1.

Groenendijk, Jeroen and Martin Stokhof (1984). "Studies on the semantics of questions and the pragmatics of answers". PhD thesis. Institute for Logic, Language & Computation (ILLC).

Heim, Irene (1983). "On the projection problem for presuppositions". In: *Formal semantics–the essential readings*, pp. 249–260.

Kao, Justine T et al. (2014). "Nonliteral understanding of number words". In: *Proceedings of the National Academy of Sciences* 111.33, pp. 12002–12007.

Qing, Ciyang, Noah D Goodman, and Daniel Lassiter (2016). "A rational speech-act model of projective content." In: *CogSci*.

Roberts, Craige (1996/2012). "Information structure: Towards an integrated formal theory of pragmatics". In: *Semantics and Pragmatics* 5, pp. 6–1.

Schlenker, Philippe (2009). "Local contexts". In: *Semantics and pragmatics* 2, pp. 3–1.

— (2021). "Triggering presuppositions". In: *Glossa: a journal of general linguistics* 6.1.

Simons, Mandy (2001). "On the Conversational Basis of Some Presuppositions". In: *Semantics and Linguistic Theory*. Vol. 11, pp. 431–448.

Simons, Mandy, David Beaver, et al. (2017). "The best question: Explaining the projection behavior of factives". In: *Discourse processes* 54.3, pp. 187–206.

Simons, Mandy, Judith Tonhauser, et al. (2010). "What projects and why". In: *Semantics and linguistic theory*. Vol. 20, pp. 309–327.

Stalnaker, Robert (1974). "Pragmatic presuppositions". In: *Semantics and Philosophy*. Ed. by Milton K. Munitz and Peter K. Unger. New York University Press, pp. 135–148.

— (1978). "Assertion". In: *Formal Semantics: The Essential Readings*, pp. 147–161.

Van Kuppevelt, Jan (1995). "Discourse structure, topicality and questioning". In: *Journal of linguistics* 31.1, pp. 109–147.

van Rooy, Robert (2003). "Questioning to resolve decision problems". In: *Linguistics and Philosophy* 26.6, pp. 727–763.

van Rooy, Robert and Katrin Schulz (2004). "Exhaustive interpretation of complex sentences". In: *Journal of logic, language and information* 13.4, pp. 491–519.

Appendix

	QUD_1 Need visa?	QUD_2 Free drink?	u_1 US	u_2 not US	u_3 GC	u_4 not GC	u_5 ϵ
w_1 (US citizen)	no	no	1	0	0	1	1
w_2 (Green card holder)	no	yes	0	1	1	0	1
w_3 (None of the above)	yes	no	0	1	0	0	1

Table 1: Worlds, utterances, and QUDs in the RSA model of example (1-a).

	prior	u_1 O.S.	u_2 Not O.S.	u_3 Runner	u_4 not runner	u_5 Athlete	u_6 Not athlete	u_7 ϵ
w_1 (Ol. Spr.)	1%	1	0	1	0	1	0	1
w_2(Runner (other))	5%	0	1	1	0	1	0	1
w_3(Athlete (other))	10%	0	1	0	1	1	0	1
w_4 (Other)	84%	0	1	0	1	0	1	1

Table 2: Worlds, utterances, and priors in the RSA model for family-genus-species inferences.