Non-maximality effects in gestural plural predication

Stavroula Alexandropoulou
University College London
London, UK
stavroulaale@gmail.com

Kurt Erbach
Goethe-University Frankfurt am Main, Germany
& Saarland University, Saarbrücken, Germany
erbach@hhu.de

Richard Breheny
University College London
London, UK
r.breheny@ucl.ac.uk

Clemens Mayr
Georg-Austin-Universität Göttingen
Göttingen, Germany
clemens.steiner-mayr1@uni-goettingen.de

Jacopo Romoli
Heinrich-Heine-Universität
Düsseldorf, Germany
jacopo.romoli@gmail.com

Yasutada Sudo
University College London
London, UK
y.sudo@ucl.ac.uk

Abstract

The present paper investigates homogeneity and non-maximality of novel pro-speech gestures. Our experiment adopts and refines Tieu, Schlenker, and Chemla's (2019) experimental design and also incorporate Romoli et al.'s (2024) idea of context manipulation. The results indicate not only that homogeneity inferences can arise from unfamiliar gestures but also that non-maximality triggered by gestural means is contextually modulated and is more robustly so in positive than in negative sentences. In light of Romoli et al.'s (2024) observation for plural definites, these findings point to the possibility that the mechanism responsible for homogeneity and non-maximality extends beyond conventionalized linguistic expressions to include unfamiliar gestures, aligning with the view that homogeneity and non-maximality emerge from productive semantic and pragmatic processes.

1 Introduction

There has recently been growing interest in the meaning contributions of gestures (Schlenker 2019; Hinterwimmer, Patil, and Ebert 2021; Ebert 2024, *inter alia*). One of the central questions in this literature is whether or not certain types of meaning commonly observed with linguistic expressions can arise from utterances containing unfamiliar gestures. Tieu, Schlenker, and Chemla 2019, among others, provide experimental evidence that scalar implicatures, presuppositions, supplements, and homogeneity can indeed be triggered by novel gestures used in place of linguistic expressions, or 'pro-speech gestures'. This paper presents further evidence that homogeneity inferences can be triggered by unfamiliar pro-speech gestures and furthermore that they give rise to non-maximality effects similarly to definite plurals.

1.1 Definite plurals

Both homogeneity and non-maximality are typically discussed with respect to plural definites, although a number of other linguistic expressions also exhibit these two properties (Krifka 1996; Löbner 2000; Križ 2015; Križ 2019, *inter alia*). By way of illustration, consider the positive sentence in (1) and the negative sentence in (2).

(1) Sam took the marbles.

(2) Sam did not take the marbles.

Homogeneity has to do with a truth-value gap. Intuitively, (1) is true iff $\lambda x.$*Sam took* x applies to each relevant marble ('Sam took all of the marbles'), while (2) is true iff $\lambda x.$*Sam didn't take* x applies to every individual marble ('Sam didn't take any of the marbles'). Note that (2) is stronger than the logical negation of (1) ('It is not true that Sam took every marble'). Therefore, there are situations where neither of these sentences are true, namely, when Sam took some but not all of the marbles. In other words, homogeneity has to do with the inference that the atomic parts of the plural definite behave homogeneously—all true or all false—with respect to the predicate (e.g. $\lambda x.$*Sam didn't take* x).

Non-maximality, on the other hand, has to do with looseness in meaning. In fact, the above characterizations of the truth-conditions of (1) and (2) are arguably too strong, since, as a matter of fact, these sentences may be judged to be true, relative to certain nonhomogeneous scenarios. Concretely, (1) can well be judged to be true when Sam took nine out of ten marbles, at least if a small number of exceptions are somehow deemed insignificant and are tolerated.

Several competing theories have been proposed to capture these two interpretive properties of plural definites. One dimension along which the theories diverge in predictions is whether the context-sensitivity of non-maximality is asymmetric with respect to polarity such that negative sentences like (2) are less context sensitive with respect to exception tolerance (e.g., Bar-Lev 2018; Bar-Lev 2021), or symmetric (Križ 2015; Križ 2016). Recently, Romoli et al. 2024 experimentally tested these predictions with a picture-sentence verification task with context. They compared an existential context ("Opening the presents is prohibited") and a universal context ("Opening the presents is required"), and found an asymmetric interpretation pattern between definite plurals in positive (e.g., *Every girl opened her presents*) and negative sentences (e.g., *No girl opened her presents*) such that context had a more robust effect on positive sentences than on negative sentences. We incorporated their design in the experiment we will report below.

1.2 Gestures

Turning to gestural semantics, Tieu, Schlenker, and Chemla 2019 provide experimental evidence suggesting that unfamiliar pro-speech gestures can be interpreted with homogeneity. However, we think that their conclusion is not entirely convincing due to a certain feature of their experimental design, which we will discuss immediately below. Certain design considerations of our own experiment are related to this issue, so let us delve into the relevant aspect of their study.

Tieu, Schlenker, and Chemla 2019 tested two utterances with the same pro-speech gesture. In both cases, a narrator in a video first introduced a background scenario with two sets of objects—three crosses introduced on the left and three coins introduced on the right—using gestural plurals in the form of repeated pro-speech gestures at different adjacent spatial points, as in (3). "CROSS-REP3" and "COIN-REP3" are the pro-speech gestures and everything else was linguistically uttered.

(3) Sam is participating in a treasure hunt in the forest, and she is looking for crosses and coins. Very quickly, Sam will find $[\text{CROSS-REP3}]_{\text{left}}$ and $[\text{COIN-REP3}]_{\text{right}}$.

Next, the narrator made an utterance beginning either with "Sam will" or "Sam will not" followed by a pro-speech two-handed grabbing gesture, TAKE-2-HANDED, performed on the right-hand side, where the coins had been introduced. The question is whether this gesture is interpreted homogeneously similarly to 'take the coins' or not.

After seeing the target utterance with a pro-speech gesture, the participant did an inference task where they indicated on a slider how strongly they thought a written inference followed from the utterance. The positive utterance was paired with the inference *Sam will take all of*

Approach	Prediction	Context	
		Existential	Universal
Asymmetric	Sam will TAKE-2-HANDED.	✓	×
	Sam will not TAKE-2-HANDED.	×	×
Symmetric	Sam will TAKE-2-HANDED.	✓	×
	Sam will not TAKE-2-HANDED.	×	✓

Table 1: Predictions of two approaches for target sentences with respect to non-maximality.

the coins and the negative utterance was paired with the inference *Sam will not take any coins.*

The results indicate that for both positive and negative sentences, participants tended to judge these inferences to follow (relative to a baseline condition where the target inference does not follow). These results led Tieu, Schlenker, and Chemla to conclude that homogeneity can be triggered by unfamiliar gestures.

While their results are suggestive, we think they do not provide a very convincing argument for their conclusion. This has to do with the nature of the inference task that they employed. The inference task certainly has a number of practical advantages, but it also has some potential pitfalls. Above all, it does not make it clear whether the inference drawn is due to the semantic or pragmatic meanings of the expressions used, or alternatively, simply due to mere plausibility arising from the context. This latter possibility is highly relevant in interpreting Tieu, Schlenker, and Chemla's (2019) experimental results. That is, we think it is not too far-fetched to suppose that participants tended to assume, even without the target utterance, that Sam would take all of the coins, if he took any of them, because, after all, coins are valuable, and explicitly described as such, and Sam was looking for them. Our experimental design is free from such a contextual inference.

2 Experiment

Building upon Romoli et al.'s (2024) and Tieu, Schlenker, and Chemla's (2019) studies, we set out to investigate homogeneity and non-maximality with pro-speech gestures. Our experiment also aims to evaluate the predictions made by the symmetric and asymmetric approaches to the context-sensitivity of non-maximality. Table 1 summarizes the predictions of the two approaches.

2.1 Methods

We employed a picture-video judgment task. At the beginning of the experiment, a written background context introduced the main character, Sam, who cleans rooms and dusts shelves.

Following Romoli et al. 2024, we manipulated contexts between participants in terms of families with different cleaning instructions. In the Existential condition, the family instructions included the statement "Taking anything off the shelves is prohibited", and in the Universal condition, "Taking everything off the shelves is required". These instructions were mentioned as part of the instructions and also displayed at the top of the screen in each trial (see Figure 1).

In each trial, a narrator named Kurt in the video (played by the second author) verbally introduces the shelves Sam is going to dust along with the objects on them, using deictic co-speech gestures directed at the shelves and objects. The narrator then makes a prediction about what Sam will do while dusting the shelves. The narrator's prediction consists of a mixed utterance similar to those in Tieu, Schlenker, and Chemla 2019 such as (4).

(4) Sam will/will not TAKE-2-HANDED_  .

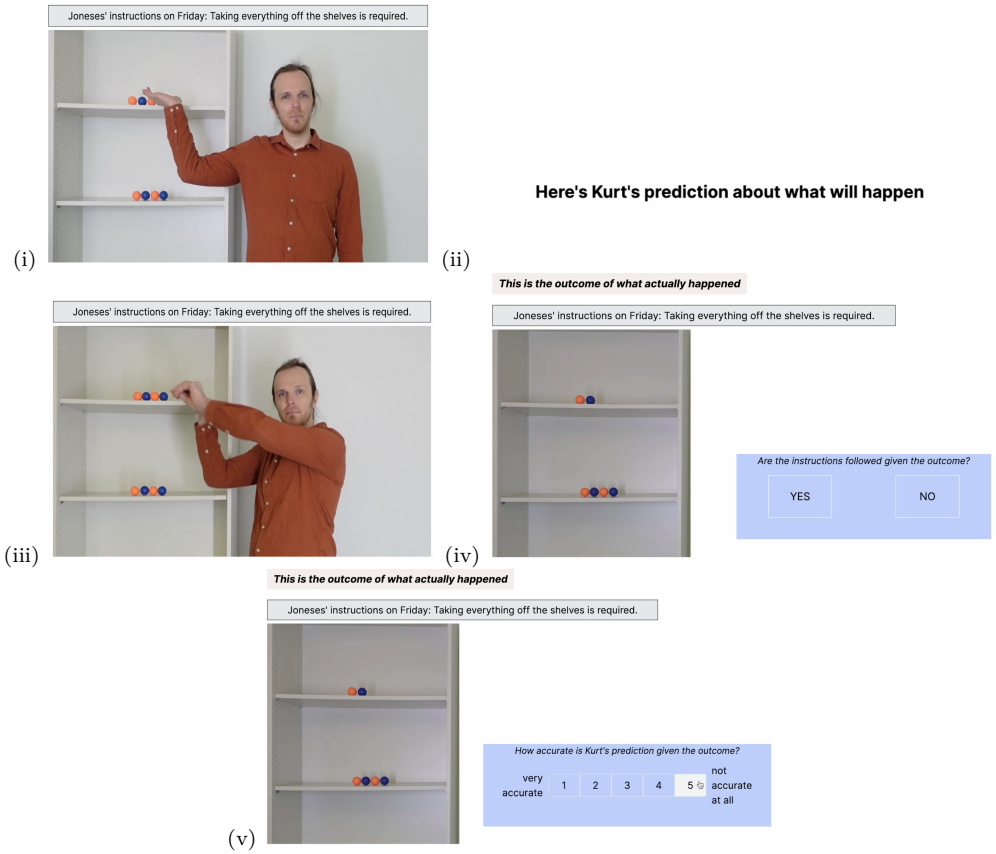

Figure 1: Example images of the displays in a trial: (i) The first display shows the video of the narrator presenting the shelves Sam will dust and the objects on them; (ii) the second display prepares participants for the narrator's prediction, containing the critical utterance; (iii) the third display shows the video with the narrator's prediction along with the critical gesture; (iv) the fourth display includes the secondary task; and (v) the fifth display contains the main task.

Unlike Tieu, Schlenker, and Chemla's (2019) items, which were about taking coins, we used objects of no particular value in all our items, in order to circumvent the potential issue of the contextual inference we pointed out above.

After the narrator's prediction, an image is presented depicting the outcome of what Sam actually did. Participants had to judge on a 5-point Likert scale how accurate the narrator's prediction was given the outcome of what Sam actually did. Before making this judgment, participants also answered a secondary question about whether Sam had followed the cleaning instructions (appearing at the top of the screen throughout the trial) given the actual outcome by *Yes* or *No*. This ensured that participants understood the context they were in and remembered it throughout the experiment. Each trial consisted of 5 displays (see Figure 1).

We manipulated the polarity of the narrator's prediction (positive: *will* vs. negative: *will not*), the type of outcome (True Control, False Control, Target), and the type of cleaning instructions (Context). In the Target outcome condition (see examples on OSF), the outcome picture showed two out of four blocks/balls, creating a non-homogeneous scenario that would only be true with non-maximality. The outcome pictures in the Control conditions depicted homogeneous scenarios, showing either all objects removed from the affected shelf or all objects intact (see OSF). The Control conditions tested homogeneity. Polarity and Type of outcome were within-participant factors and Context was a between-participant factor.

We created two experimental items: one featuring 4 balls on each shelf and the other featuring 4 blocks on each shelf. The position of the balls/blocks affected by the narrator's

gesture (top/bottom shelf), the direction of the gesture (left/right) and the narrator's outfit (3 different outfits) were counterbalanced across 12 lists. In half of them the order of labels of the response buttons (see Figure 1) was presented in one sequence, while in the other half, the order was reversed. In total, there were 24 lists—12 for each Context. Each item appeared in 6 conditions (2 Polarity × 3 Outcome type) per list, resulting in a total of 12 experimental trials per list. We also created 24 filler items, each featuring different types of objects (candles, mugs, toys) on both shelves and various gestures (e.g., a one- or two-handed pick-up gesture affecting different shelves, a one- or two-handed grabbing gesture affecting different shelves, a one-handed moving up/down gesture, and a one-handed knocking-down gesture). The same 24 filler items appeared in every list and were designed to balance the yes/no responses to the secondary question in experimental items regarding whether the cleaning instructions were followed by Sam, achieving a near-optimal balance.

Participants were randomly assigned to one of the two Context conditions. After reading the instructions, participants were presented with two example items designed to familiarize them with the tasks. For each example item, they received explicit feedback on the correct responses to the secondary task—whether Sam followed the cleaning instructions—and the critical task, if they initially gave an incorrect response. Unambiguous items resembling the filler items were used as example items. Participants could only proceed to the main experimental session after providing the correct responses to these example items.

The experiment was implemented in Gorilla Experiment Builder. Native speakers of English were recruited online via Prolific. To achieve a target of 40 participants per list, we recruited a total of 960 participants across two experimental batches. Each participant was compensated £2.61 (1st batch) or £2.55 (2nd batch) for their participation. Informed consent was obtained from each participant prior to participation. We excluded participants who completed the study on a smartphone, or on the Safari browser, which is not optimal for Gorilla experiments involving images and videos. Additionally, participants who reported not hearing any sound or who had an accuracy rate below 75% on the critical task on the Control items (with ratings higher than 3 for False Controls and ratings lower than 3 for True Controls considered errors) were excluded, resulting in a final dataset of 862 participants. This set of participants had a mean age of 39.94 years (age range: 18–77). Of these, 490 identified as female, 368 as male, and 4 preferred not to say.

2.2 Results

Following Romoli et al. 2024, the Context (Type of Instructions) factor was recoded into Lax (positive predictions in existential context, negative predictions in universal context) and Strict (positive predictions in universal context, negative predictions in existential context). Lax invites non-maximal interpretations while Strict prevents them. Figure 2 shows mean ratings per condition. Notice the high mean ratings obtained by the True control conditions overall, suggesting the availability of homogeneity effects for both positive and negative sentences. A similar, but inverse, pattern is observed for the False control conditions.

We fitted a cumulative link mixed-effects model to analyze prediction ratings for Target items, with sum-coding for Context and treatment-coding for Polarity of Prediction (with negative as the reference category).[1] The analysis was conducted using the ordinal package (Christensen 2023) in R (R Core Team 2022). The model also included random intercepts for Participants and random by-Participant slopes for Polarity. Due to the item factor having only two levels, the variance attributed to it was unlikely to be meaningful, and we excluded it as a random effect from our model.

[1]The model including the True and False Control conditions failed to converge, potentially due to insufficient variability in the corresponding ratings.

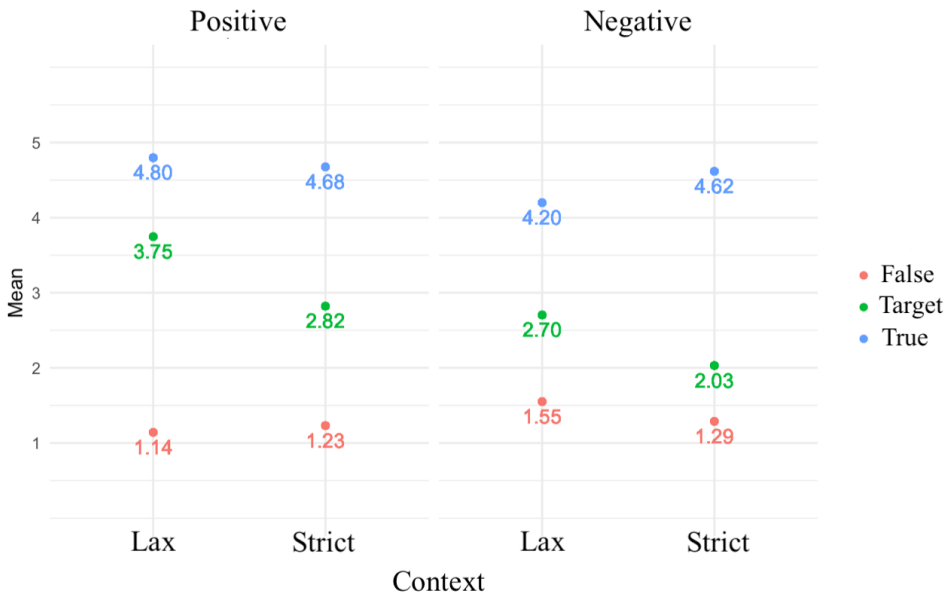

Figure 2: Mean prediction ratings per condition.

Our analysis revealed (see whole output in the Appendix on OSF): (i) a strong positive effect of Polarity of Prediction ($\beta = 2.25$, $SE = 0.13$, $z = 17.11$, $p < 0.0001$), with positive predictions eliciting higher ratings overall compared to negative predictions, (ii) a strong positive effect of Context ($\beta = 1.75$, $SE = 0.16$, $z = 10.64$, $p < 0.0001$), with Lax conditions receiving higher ratings than Strict ones overall, and importantly (iii) a significant positive interaction effect ($\beta = 0.46$, $SE = 0.19$, $z = 2.41, p < 0.05$), indicating greater context-sensitivity for non-maximal interpretations in positive predictions compared to negative ones.

3 Discussion and conclusion

The intermediate judgments in the Target conditions, which involved non-homogeneous scenarios, suggest that pro-speech gestures do trigger homogeneity.

Our results furthermore indicate that non-maximality of pro-speech gestures is more context-sensitive in positive sentences than in negative sentences, similarly to what Romoli et al. 2024 observed for plural definites. This points to the possibility that the mechanism responsible for homogeneity and non-maximality inferences may extend beyond linguistic expressions to include gestures. Following Schlenker 2019, the potential for gradient iconic effects of the pro-speech gesture tested in the present study makes it unlikely that the gesture was directly translated into the corresponding linguistic item.

Overall, our findings involving unfamiliar gestures further support the idea that types of inferences typically observed with linguistic expressions emerge through productive semantic and pragmatic processes rather than being exclusively tied to language.

Acknowledgements. The authors wish to thank Cornelia Ebert and Markus Steinbach for their valuable feedback on the experiment. The research in this paper was supported by an AHRC-DFG grant (AHRC: AH/V003526/1; DFG: EB 523/2-1, STE 958/12-1, STE 2555/3-1).

References

Bar-Lev, Moshe (2018). "Free choice, homogeneity and innocent inclusion". PhD thesis. The Hebrew University of Jerusalem.

Bar-Lev, Moshe E. (2021). "An Implicature Account of Homogeneity and Non-Maximality". In: *Linguistics and Philosophy* 44.5, pp. 1045–1097. DOI: 10.1007/s10988-020-09308-5.

Christensen, Rune H. B. (2023). *ordinal—Regression Models for Ordinal Data*. R package version 2023.12-4.1. URL: https://CRAN.R-project.org/package=ordinal.

Ebert, Cornelia (2024). "Semantics of gesture". In: *Annual Review of Linguistics* 10.1, pp. 169–189.

Hinterwimmer, Stefan, Umesh Patil, and Cornelia Ebert (2021). "On the interaction of gestural and linguistic perspective taking". In: *Frontiers in communication* 6, p. 625757.

Krifka, Manfred (1996). "Pragmatic Strengthening in Plural Predications and Donkey Sentences". In: *Proceedings of SALT VI*. Ed. by Teresa Galloway and Justin Spence. Ithaca, NY: CLC Publications, pp. 136–153.

Križ, Manuel (2015). "Aspects of homogeneity in the semantics of natural language". PhD thesis. University of Vienna.

— (2016). "Homogeneity, maximality, and *all*". In: *Journal of Semantics* 33, pp. 493–539.

— (2019). "Homogeneity effects in natural language semantics". In: *Language and Linguistics Compass* 13.11, e12350. DOI: 10.1111/lnc3.12350.

Löbner, Sebastian (2000). "Polarity in natural language: Predication, quantification and negation in particular and characterizing sentences". In: *Linguistics and Philosophy* 23.3, pp. 213–308.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Romoli, Jacopo et al. (2024). *Putting plural definites into context*. submitted.

Schlenker, Philippe (2019). "Gestural semantics: Replicating the typology of linguistic inferences with pro-and post-speech gestures". In: *Natural Language & Linguistic Theory* 37, pp. 735–784.

Tieu, Lyn, Philippe Schlenker, and Emmanuel Chemla (2019). "Linguistic inferences without words". In: *Proceedings of the National Academy of Sciences* 116.20, pp. 9796–9801.