

Experimenting with the King of France

Márta Abrusán & Kriszta Szendrői

Lichtenberg Kolleg, Göttingen, Germany & University College London, UK
 abrusan@alum.mit.edu & k.szendroi@ucl.ac.uk

Abstract. Definite descriptions with reference failure have been argued to give rise to different truth-value intuitions depending on the local linguistic context in which they appear. We conducted an experiment to investigate these alleged differences. We have found that pragmatic strategies dependent on verification and topicalization, suggested in the context of trivalent theories, indeed play a role in people's subjective judgments. Overall however we think that our findings are best explained by combining these pragmatic strategies with an approach that assumes that definite descriptions have a bivalent semantics, as well as a pragmatic presupposition attached to them. We also suggest that the verification of a sentence –where possible – proceeds through a *pivot* constituent, and that this concept is relevant for the proper description of how speakers understand semantic meaning.

Keywords: definite descriptions, presupposition, topics, verifiability, experimental pragmatics

1 Introduction

According to Russell (1905, 1957), definite descriptions assert the existence of a unique individual that satisfies the description. When such an individual does not exist, as in (1), the sentence makes a false assertion. Famously, Strawson (1950, 1964) argued that Russell's theory cannot predict why speakers (like him) feel "squeamish" about assigning the truth-value 'false' to sentences such as (1). His proposal (which can be traced back to Frege 1892) was that definite descriptions instead of asserting, presuppose the existence of a unique individual that satisfies the description: when this presupposition is not met, the question of truth or falsity does not even arise and so the sentence does not have a truth-value. A third possibility (cf. Stalnaker 1974, 1978) is that definite descriptions both assert and presuppose the existence of a unique individual satisfying the description.

- (1) The king of France is bald.

Strawson (1964) has noted however that truth-value intuitions change when the same definite description is put in different contexts. He conceded that according to his intuition examples such as (2) do appear to be straightforwardly false. This in turn

requires some explanation from the view according to which (2) should not have a truth-value.

- (2) The exhibition was visited yesterday by the king of France.

Since Strawson's observation, various factors have been identified in the literature that might influence our truth-value intuitions about sentences with reference failure noun phrases. Strawson (1964) himself (cf. also Reinhart 1981 and many others) have identified topic-comment structure as a factor. Others (cf. Fodor 1979, Lasnik 1993, von Stechow 2004) stressed the importance of background knowledge based on which the sentence could be verified (or not), independently of the existence of the problematic referent. Sometimes, conflicting judgments have been asserted, which makes it hard to judge the relative import of the theories.

We have conducted a behavioral experiment that was designed to capture people's intuitions about sentences like (1)-(2). Our overall results argue against Russell's non-presuppositional semantics for definite descriptions and also against Strawson's truth-value gap approach, and suggest that the pragmatic presupposition approach of Stalnaker is more likely to be correct. But we have also found that the above-mentioned Strawsonian theories all identify relevant pragmatic factors that influence truth-value intuitions, even though none of them can fully predict the behavioral pattern we found: The sentence's topic-comment structure (as suggested by Strawson 1964, Reinhart 1981, etc) and the sentence's verifiability (as suggested by Lasnik 1993, von Stechow 2004) are *both* important factors influencing truth-value judgments. At the end of this paper, we offer a brief discussion of how to reconcile these seemingly divergent factors.

2 Possible Factors Behind Wavering Truth-Value Judgments

Strawson (1969) has proposed that one factor behind difference in (1) and (2) is the topic-comment structure of the sentence. Topics are understood to be the constituents that the sentence is pragmatically about. Strawson proposed that when the definite description is not in a topic position, it is "absorbed" into the meaning of the predicate, and since it is not a referring expression anymore its presupposition is turned into an existential statement. This predicts the difference between (1) and (2): The noun phrase the king of France is in topic position in (1) hence in a context where it is known that France has no king, it leads to a presupposition failure, associated with "squeamishness". In (2) the same noun phrase is not in a topic position and so the sentence is simply false. The topic approach has been adopted and developed further by many researchers, most importantly Reinhart (1981, 1995). (See also Hajicova 1984, Gundel 1977, Lambrecht 1994, Erteschik-Shir 1997, Atlas 2004, Geurts 2007, Shoubye 2009, among others).

However, the importance of topichood has been questioned (cf. von Stechow 2004). One reason to doubt that the existential presupposition is absorbed into the predicate (or is just non-existent) when the definite description is not in topic is that

the existential presupposition of definites seems to project out of embedded contexts, such as the antecedents of conditionals, whether or not the definite is in topic position. Thus (3) still seems to imply the existence of a French king, which suggests that the definite is still presuppositional, despite not being in topic.

- (3) If the exhibition was visited by the king of France, the organizers must be happy.

This latter observation highlights a very important point, identified most clearly by von Fintel. Namely, that our intuitions about accepting or rejecting a sentence as true or false, and the sentence's actual semantic truth-value (and hence its presuppositionality) are two separate things. Speakers might feel that a sentence is false or true even when semantically it has no truth-value, as long as they can find some reason based on which they can reject (or accept) the sentence. The feeling of “squeamishness” arises only when all pragmatic repair strategies for dealing with a truth-valueless sentence fail.

The first proposal in this spirit was due to Lasnik (1993). His main focus were examples such as (4) which are said to be judged false. Lasnik's observation was that in a situation where the chair in front of the speaker is empty, or when it is occupied by somebody other than the king of France, speakers have enough grounds to reject the sentence as false: They can look at the chair and see that the king of France (whether or not he exists) is not in it.

- (4) The king of France is sitting in this chair.

In the case of (1), in the absence of background knowledge about the hairstyle of French royals, speakers do not have enough grounds to reject or accept the sentence, and are left with the feeling of squeamishness. Lasnik's theory rests on the following tenets: **(a)** Assume two kinds of truth-values: (i) semantically assigned values 1, 0, and a third value #, which corresponds to ‘neither 0 or 1’ (ii) pragmatically assigned values TRUE, FALSE which represent the status of a sentence with respect to a given body of information, and correspond to acceptance and rejection. **(b)** Once we are faced with presupposition failure (neither 0 nor 1), there are fall-back strategies to fill in the gap and arrive at TRUE and FALSE. **(c)** Lasnik's fall-back strategy: Step 1: revise the given body of information to remove the knowledge that there is no king of France. Step 2: See if the given body of information can be consistently extended to include the target proposition.

A consequence of Lasnik's analysis is that only those propositions are predicted to have the truth-values TRUE or FALSE that are in direct conflict (or in accordance) with what can be concluded from the given body of information. This conclusion has been argued to be too weak by von Steinhilber (2004), based on examples such as (5), which he argues is felt to be FALSE, even in the absence of any information about who is on a state visit in Australia this week.

- (5) The king of France is on a state visit to Australia this week.

He proposes to add another fall-back strategy, besides (c) above:¹² (d) Rejection/acceptance might (also) be based on the possibility of examining the intrinsic properties of a contextually salient independent entity (that everyone agrees exists).

This suggestion explains (5), even if the given body of information does not contain anything about who was visiting Australia. In principle, we could examine the properties of Australia and see whether the king of France is in it or not. *Australia* thus serves as a salient foothold for verification, based on which the truth of the sentence in (5) can be evaluated.

3 The Experiment

We tested 33 native speakers of English (mostly British English, all of them familiar with basic elements of British culture), aged 20-55, most of whom participated for a small fee. We investigated how participants judge different types of sentences with reference failure noun phrases. The participants first read instructions given to them on the computer screen, reproduced in (6).

- (6) *In this experiment, statements will appear on your screen. If you think a statement is true, you should click on the 'TRUE' button. If you think a statement is false, you should click on the 'FALSE' button. Sometimes, it may happen that you cannot decide. In those cases, you should click on the 'CAN'T SAY' button. Please do not dwell on your decision for too long. There is no right or wrong answer!*

After a short practice session, participants were left alone with a program which presented the test items one-by-one on the screen. Each item contained one sentence, as shown in (7). Participants could use the mouse to click on the buttons. After they chose an answer, the next item appeared automatically.

- (7) *Example of an experimental trial:*

The king of France is bald.		
FALSE	CAN'T SAY	TRUE

There were eleven test conditions, with eight test items in each condition. The test items were obtained by placing 8 definite descriptions that lack referents (listed in (8))

¹ That both (c) and (d) are needed is suggested by examples that fall under (c) but not (d): *The king of France can jump 100 feet into the air unaided*. (Example from intro by Bezuidenhout and Reimer to von Fintel 2004). However, if independent footholds for rejection/verification can also be general laws, then Lasersohn's account can be subsumed under von Fintel's, and thus (c) is a subcase of (d). It seems to us however that this would make it very hard to track what predictions von Fintel's proposal actually makes.

in eleven types of sentential contexts, (the test conditions), illustrated in (9). More on the test conditions below.

- (8) the king of France; the emperor of Canada; the Pope's wife; Princess Diana's daughter; the beaches of Birmingham; the Belgian rainforest; the coral reefs of Brighton; the volcanoes of Kent.
- (9) Examples of test conditions, illustrated here with *the king of France*
 - 0 The king of France is bald.
 - 1 France has a king and he is bald.
 - 2 The king of France is on a state visit to Australia this week.
 - 3 The king of France is married to Carla Bruni.
 - 4 The king of France, he was invited to have dinner with Sarkozy.
 - 5 Sarkozy, he was invited to have dinner with the king of France
 - 6 The king of France isn't bald.
 - 7 The king of France is not on a state visit to Australia this week.
 - 8 The king of France is not married to Carla Bruni.
 - 9 The king of France, he wasn't invited to have dinner with Sarkozy.
 - 10 Sarkozy, he wasn't invited to have dinner with the king of France.

The 88 test items were supplemented by almost twice as many filler items containing true and false controls. Altogether, there were 253 items presented in three blocks. The items were pseudo-randomised: there were no items from the same condition, or with the same NP, closer than 4 trials. The statements were presented in three separate blocks.

All three theories agree that sentences like *The king of France is bald*, i.e. Condition 0, should lead to truth-value gaps and predict FALSE judgments for the non presuppositional assertions in Condition 1. The theory of von Steinhilber predicts a higher number of FALSE judgments in Condition 2, 3, 4, 5 than in Condition 0 because the former contain an independent NP alongside the referentially challenged NP. It also does not predict any difference among Conditions 2, 3, 4, 5. Lasnik predicts a higher percentage of FALSE answers for Condition 3: The items in this condition were individually paired with 8 true control items, e.g. the control item for (12-3) and (12-8) was that Sarkozy is married to Carla Bruni. It was predicted that participants who judge control items TRUE would have the necessary knowledge to reject the corresponding items in Condition 3. Strawson and Reinhart predicted that items in Condition 4 lead to truth-value gaps because in these the referentially challenged NP is topicalized, while their variants in Condition 5 in which the same NP is not in topic are judged FALSE.

Following pilot studies, we have discovered that a large proportion of our participants judged items in our base condition, Condition 0, as FALSE. For this reason, we included a further five conditions, Conditions 6-10, which corresponded to Conditions 0, 2-5, respectively, but which involved sentential negation. Our assumption was that a FALSE judgment that reflects a rejection of a particular statement based on it being semantically false should turn into a TRUE judgment once the statement in question

is negated. If a negated sentence such as *The king of France is not bald* is not judged to be TRUE, then likely the corresponding positive sentence is judged to be FALSE for some other reason than having the semantic value 0. Further, if *The king of France is not bald* is not judged to be TRUE, the non-Russellians are still in business and we can test the validity of the theories presented above. In particular, we make the following predictions. We would get a higher number of TRUE judgments in Condition 7, the negated version of Condition 2, compared to Condition 6, the negated version of Condition 0, to support von Fintel’s theory. Lasersohn predicted a higher number of TRUE judgments in Condition 8, the negated version of Condition 3, than in Condition 6. While Reinhart/Strawson predicted higher number of TRUE judgments in Condition 10 than in Condition 9, which are the negated equivalents of Conditions 6 and 5 respectively.

4 Results

4.1 Data

All but one of our participants successfully finished the task. This participant was excluded. We compared the proportion (%) at which subjects replied FALSE to the test items in Conditions 0-5 with ANOVA, following checks for normal distribution and homogeneity of variance. We found only weak or nearly significant differences between any of the conditions 0-5 ($p > 0.05$, Bonferroni post-hoc test), i.e. the positive conditions. Our subjects said FALSE to most of these most of the time. This was counter to our predictions, according to which we expected a low number of FALSE judgments in Condition 0, and a high number of FALSE judgments in Condition 1. Neither of the conditions 0-5 differed significantly from our FALSE-controls either.

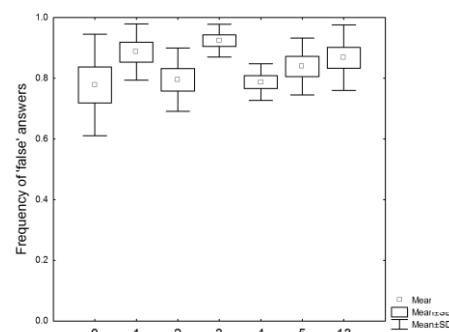


Fig. 1. Frequencies of FALSE answers for conditions 0-5, and the control condition 13 for which we expected FALSE answers.

Since there was no significant difference in the percentage of FALSE judgments between any of the conditions 0-5, we could not verify either a von Fintel effect (0 vs. 2), nor a Lasersohn effect (0 vs. 3) nor a topic-effect (4 vs. 5) by looking at positive sentences.

The negated versions of Conditions 0-5 (except Condition 1), namely Conditions 6-10 turned out to be more informative. We compared the proportion (%) of TRUE responses in these conditions with ANOVA, (Bonferroni post-hoc test). As

Figure 2 illustrates, we found that speakers responded TRUE at a significantly higher proportion to the items in Conditions 7, 8, 9, 10 than to the items in Condition 6 ($p < 0.003$ in all cases), our base-line. The significant difference between Conditions 6 and 7 (and also 6 vs. 8, 9, 10) indicates that there is a von Fintel-effect. The significant difference between Conditions 6 and 8 supports Lasersohn's theory. Condition 8 also differed significantly from Condition 7 ($p = 0.0012$). In addition, we also found a significant difference ($p = 0.037$) between Conditions 9 vs. 10: speakers responded with TRUE at a significantly higher proportion to Condition 10 than to Condition 9, which is in accordance with Strawson/ Reinhart's predictions, and indicates that topichood also plays a part in subjective evaluations of the truth of sentences with reference failure definite descriptions. The difference between conditions 7 vs. 9 and 8 vs. 10 was not significant ($p > 0.05$).

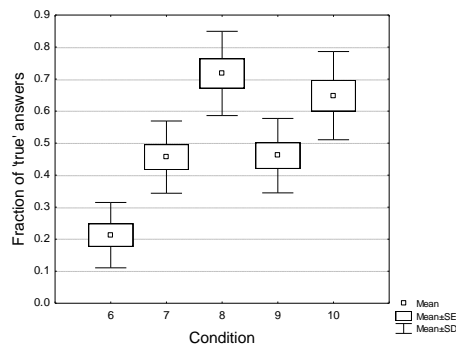


Fig. 2. Frequencies of TRUE answers for conditions 6-10. Box represents standard errors of the sample, whiskers standard deviation.

When comparing Conditions 6-10 with true controls, we found that all the conditions 6-10 (except condition 8) differed significantly from true controls as well, even if weakly ($p < 0.045$). It is hard however to interpret this result, as Conditions 6-10 contained negation, but our control sentences did not.

4.2 Interpreting the Results

Positive vs. negative conditions. We found that the responses to Condition 0 did not differ significantly from Condition 1, or from our control condition involving non-problematic false statements. At a first glance, this result might seem to support the Russellian view. However, in the light of the answers we got for Conditions 6-10, the negative conditions, the Russellian position is hard to maintain as this view has no room to predict the systematic variation in conditions 6 to 10.

The truth-value gap account of Strawson (also assumed by Lasersohn 1993 and von Fintel 2001) at first sight seems to have more room for maneuvers that might explain the overall pattern of data we found. On this account, all the sentences in conditions 0, 2-10 are truth-valueless: nevertheless it is possible to predict the results we got by assuming that our subjects interpreted the response option FALSE 'I do not think it is true' or 'I am rejecting this sentence' rather than 'I think this sentence is false'. Sentences in the positive Conditions 2-5 are judged to be FALSE either for the

same reason, or because of Lasersohn's/von Fintel's idea according to which speakers opt for FALSE in order to avoid the conversational impasse created by a presupposition failure, but only if they have independent reasons to reject the sentence. The same reasoning also predicts a difference between conditions 6 vs. 7-10, which was also confirmed by our results. The problem is when we compare our answers in Condition 0 and Condition 6: To the extent that the above approaches would go along with our explanation for the high number of FALSE judgments in Condition 0 as instances of overapplication of FALSE to truth-valueless sentences, the same reasoning should apply to Condition 6. However, a post hoc comparison of the number of FALSE results for Conditions 0 and 6 revealed a significant difference (t-test $p=0.0002$), our subjects were much more likely to reply FALSE to the examples in Condition 0 than to the examples in Condition 6. We do not see how to reconcile this finding with the truth-value gap approach.

The theory of pragmatic presuppositions, according to which definite descriptions both assert and presuppose the existence of a unique individual that satisfies the description, such as Stalnaker's (1974) theory can explain the full pattern of results (cf. also Karttunen and Peters 1979; Abbott 2000; Simons 2001; Schlenker 2008, 2009; (among others)). As for the positive sentences (Condition 0), speakers say these are false, because they are indeed semantically false, and the fact that these sentences at the same time exhibit a presupposition failure is not in itself a reason to change this judgment, at least as long as false sentences do not have to be added to the context set. As for the sentences in Condition 6 (the negative baseline sentences), although these are semantically true, speakers are reluctant to mark these sentences as such because they also manifest a presupposition failure, and therefore one should avoid adding them to the context set. Suppose TRUE implies a commitment to add the sentence to the context set: this suffices to prevent people from saying that such sentences are TRUE. But when there is a good pragmatic strategy based on which the presupposition failure can be ignored, as in Conditions 7-10, the sentences with negation (that are semantically true) will be felt as pragmatically TRUE as well. This predicts the difference we found between Condition 6 vs. conditions 7, 8, 9 10.

The von Fintel vs the Lasersohn-effect. We found a von Fintel effect: Condition 7 (and also Conditions 8, 9 and 10) received a significantly higher proportion of TRUE responses than Condition 6. At the same time, we found that there was a significant difference between the proportion of TRUE responses to Condition 7, the 'pure' von Fintel condition, and Condition 8, the one that aimed to test Lasersohn's theory. We take this result to indicate that there is indeed a separate Lasersohn and von Fintel-effect, and both effects measurably influenced truth-value judgments. For the von Fintel effect it is enough that an NP is present on the basis of which it would be possible to verify the sentence (given some appropriate knowledge dataset, say Wikipedia). For the Lasersohn effect, the properties of the NP provide a basis for verification given the speaker's actual dataset. This explains the finding that the Lasersohn effect is stronger: Our sentences that satisfied Lasersohn's criteria also satisfied von Fintel's criteria, but not the other way around.

Topic effect. We also found that there was a significantly higher number of TRUE responses in Condition 10, where an existentially sound NP was topicalised, than in Condition 9, where the NP with referential failure was topicalised. Our results are thus consistent with Strawson’s and Reinhart’s prediction.

5 Topics and Verifiability

We think that overall our findings support the position that all the three proposals discussed above were right to some degree, in that they all identified relevant factors for truth value intuitions. One question that arises is whether the above are independent factors, or whether there is some connection among them.

We believe that verification and topichood are not unrelated notions. However, there is also a fair amount of confusion around the notion of topic: sometimes it is understood as a discourse based notion (cf. Kuno 1972, Gundel 1974, Givon 1983, von Stechow 1996, Roberts *to appear*, among others) but sometimes it is (also) couched in verificationalist terms (cf. Strawson 1964, Reinhart 1981, Lambrecht 1994, Erteschik-Shir 2007, among others). However, even the latter employ tests for topichood that are discourse based. We suggest to distinguish two related but different concepts from each other: a discourse based notion of *topic*, which is what the sentence is pragmatically about, and what we call *pivot*, which is the constituent in the sentence based on which the sentence is verified, in other words, the contextually salient entity that provides the foothold for verification in von Stechow’s sense. Thus pivothood is a semantic notion that is concerned with the process of understanding a sentence. This involves – at least in some cases – the possibility of verifying the sentence. The pivot is what the sentence is semantically about, if such an entity can be found, in this verificationist sense of aboutness. It is not an obligatory element of understanding a sentence: In some cases it might not be possible to identify a unique pivot or indeed any pivot. Further, pivots do not have to be constituents of the sentence. This might be the case in examples with focus marking or clefting, where the set of alternatives provided by focus (or the background question) can serve as the pivot.

We believe speakers arrive at the truth-value intuitions TRUE and FALSE that we found as follows. For the reasons described above, we believe that the facts we found can be explained best on pragmatic presupposition theory such as Stalnaker’s. Here is how the reasoning might go: (i) Sentences with reference failure NP’s are semantically false (or true, if negated), but infelicitous because of presupposition failure. (ii) The presupposition failure can be ignored (and so the predicted infelicity can disappear) if a potential pivot that is independent from the reference failure NP is present in the sentence (\approx von Stechow). (iii) The above effect is even stronger, if the independent, referentially sound NP is topicalised (\approx Strawson/Reinhart). This latter fact is because topics are default pivots, and hearers tend to verify sentences based on the sentence’s topic. (iv) The effect is also stronger if the speaker has direct knowledge about the properties of the potential pivot, which contradicts the proposition irrespective of whether the entity denoted by the reference failure NP exists or not (\approx Lasnik).

Acknowledgements. Thanks to György Abrusán for his invaluable help with statistics. We are grateful to Nathan Klinedinst and Daniel Rothschild for insightful comments on an earlier draft, as well as Klaus Abels, Bob Borsley, Judit Gervain, Vanessa Harrar, Ad Neeleman and the audiences at the UCL LingLunch and the Departmental Colloquium at Essex University for their helpful questions and suggestions. The first author acknowledges financial support by the Mellon Foundation and the Lichtenberg Kolleg, Georg-August-Universität Göttingen. All remaining errors are our own.

Selected References

1. Erteschik-Shir, Nomi (2007). *Information structure*. Oxford University
2. von Fintel, Kai (2004). Would you believe it? The King of France is back! (Presuppositions and truth-value intuitions.) In Anne Bezuidenhout and Marga Reimer, (eds.). *Descriptions and beyond: an interdisciplinary collection of essays on definite and indefinite descriptions and other related phenomena*, Oxford University Press, 315-341.
3. Givon, Talmy (1983). *Topic continuity in discourse: An introduction*. In: Givon, T. (ed). *Topic continuity in discourse: A quantitative cross-language study*. John Benjamins: Amsterdam/Philadelphia.
4. Karttunen, Lauri and Peters, Stanley (1979). Conventional Implicatures in Montague Grammar, in Oh, C.-K. and Dineen, D. (eds.), *Syntax and Semantics 11: Presupposition*, New York: Academic Press, 1–56.
5. Kuno, Susumu (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* 3, 269-320.
6. Lasersohn, Peter (1993). Existence presuppositions and background knowledge. *Journal of Semantics* 10 (2), 113-122.
7. Lambrecht, Knud (1994). *Information structure and sentence form: A theory of topic, focus and the mental representations of discourse referents*. Cambridge University Press.
8. Reinhart, Tanya (1981). Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27, 53-94.
9. Roberts, Craige (to appear) Topics. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.) *Semantics: An International Handbook of Natural Language Meaning*. Mouton de Gruyter.
10. Russell, Bertrand (1905). On Denoting. *Mind* 14, 479–493.
11. Russell, Bertrand (1957). Mr. Strawson on Referring. *Mind* 66, 385–389.
12. Schlenker, Philippe (2008) Be articulate: A pragmatic theory of presupposition, *Theoretical Linguistics*, 34: 157–212.
13. Schoubye, Anders (2009). Descriptions, Truth-value Intuitions, and Questions. *Linguistics and Philosophy* 32 (6), 583-617.
14. Simons, Mandy (2001) On the conversational basis of some presuppositions, in Hastings, R. and Jackson, B. and Zvolensky, Z. (eds.), *Proceedings of Semantics and Linguistics Theory 11*, Ithaca, NY: CLC Publications, 431–448.
15. Stalnaker, Robert C. (1974). Pragmatic Presuppositions. In Munitz, Milton and Unger, Peter (eds.) *Semantics and Philosophy: Essays*. New York University Press.
16. Stalnaker, Robert C. (1978). Assertion. In Cole, P. (ed.) *Syntax and Semantics 9*. New York: New York Academic Press, pp. 315–32.
17. Strawson, Peter F. (1950). On Referring. *Mind* 59, 320–344.
18. Strawson, Peter F. (1964). Identifying reference and truth-values. *Theoria* 30 (2), 96-118.