DE GRUYTER

Piotr Stalmaszczyk (Ed.)

PHILOSOPHICAL INSIGHTS INTO PRAGMATICS

PHILOSOPHICAL ANALYSIS

Philosophical Insights into Pragmatics

Philosophical Analysis

Edited by
Katherine Dormandy, Rafael Hüntelmann,
Christian Kanzian, Uwe Meixner,
Richard Schantz, Erwin Tegtmeier

Volume 79

Philosophical Insights into Pragmatics

Edited by
Piotr Stalmaszczyk

DE GRUYTER

Contents

List of Contributors

Pedro Abreu
New University of Lisbon, Portugal
pedroabreu@fcsh.unl.pt

Brian Ball
Oxford University, UK
brian.ball@philosophy.ox.ac.uk

Tadeusz Ciecierski
University of Warsaw, Poland
taci@uw.edu.pl

Roberta Colonna Dahlman
Stockholm University, Sweden
roberta.colonna.dahlman@gmail.com

Chi-Hé Elder
University of East Anglia, UK
c.elder@uea.ac.uk

Leopold Hess
Radboud University Nijmegen,
The Netherlands
leopoldhess@gmail.com

Palle Leth
Stockholm University, Sweden
palle.leth@philosophy.su.se

Andrei Moldovan
Universidad de Salamanca, Spain
mandreius@usal.es

Joanna Odrowąż-Sypniewska
University of Warsaw, Poland
j.odrowaz@uw.edu.pl

Pedro Santos
FCHS, University of the Algarve & Lancog,
University of Lisbon, Portugal
pmtcs36@gmail.com

Piotr Stalmaszczyk
University of Łódź, Poland
piotr.stalmaszczyk@uni.lodz.pl

Maciej Witek
University of Szczecin, Poland
maciej.witek@usz.edu.pl

Piotr Stalmaszczyk
Philosophical Insights into Pragmatics: Introduction

1 Introduction

Contemporary research within philosophy of language and linguistics frequently leads to the pragmaticization of meaning, i.e. shifting the burden of theoretical (but also quite often experimental) analysis from semantics to pragmatics.[1] Ruth Kempson's well-known description of semantics as a "bridge discipline between linguistics and philosophy" (Kempson (1977, p. ix) would have to be modified today and would have to take into consideration pragmatics and – to reapply the metaphor – its considerable bridging potential. Also, identifying philosophy of language with semantics[2] would today require a far-fetched reanalysis of semantics. On the other hand, claims that pragmatics is the appropriate (if not the only) domain for a theory of meaning are abundant.[3]

As a result of these tendencies, the field of pragmatic investigations has considerably broadened its scope and it now encompasses a wide range of topics, thereby strongly interconnecting philosophy of language with linguistics (and other disciplines).[4]

---

1 See, for example, the papers collected in Turner (ed.) (2011), a volume significantly entitled *Making Semantics Pragmatic*, and in Stalmaszczyk (ed.) (2014), with another telling title – *Semantics and Beyond*. See also the papers in Depraetere/Salkie (eds.) (2017), and Capone/Carapezza/Lo Piparo (eds.) (2018) on recent investigations into the semantics/pragmatics interface.
2 Cf. the following statement in the opening lines of an interdisciplinary reader in semantics: "The part of philosophy known as the philosophy of language, which includes and is sometimes identified with the part known as semantics" (Caton 1971, p. 3).
3 Cf. for example Woleński (2003: 120) on pragmatics as the proper place for the study of meaning, and Turner's 'Wittgensteinian' postulate to show the "semanticist the way out of semantics" (Turner 2011, p. 14). Turner also comments on truth-conditional semantics as "thoroughly and constitutively pragmatic" (Turner 2011, p. 14, and n. 18). Most recently, Jaszczolt (2018) argues that pragmatics and philosophy "have to occupy the center stage in the study of meaning in pursuit of a new [...] paradigm" (Jaszczolt 2018, pp. 155–156).
4 For a general overview of philosophical perspectives for pragmatics, see Sbisà, Östman and Verschueren (eds.) (2011), and Capone/Lo Piparo/Carapezza (eds.) (2013). For a discussion of philosophical pragmatics as a branch of philosophy of language, see Jaszczolt (2018). For some changes in the scope of pragmatic research, see the contributions to Verschueren/Östman (eds.) (2009), and most recently Depraetere/Salkie (eds.) (2017), especially Part Three ("Exploring New

Papers collected in this volume focus on several more recent topics of research, such as negotiating what is said, analyses of game-playing situations, subsentential speech acts, accommodation in linguistic interaction, expressive meanings and expressive commitments, and also on some well-established concepts (such as presupposition, entailment, implicature, speech acts, different cases of meaning as use) viewed from new perspectives. It is hoped that the discussed topics contribute to the field of philosophical pragmatics, regarded as a branch of philosophy of language (in the sense of Jaszczolt 2018).

2 Contents of the volume

Roberta Colonna Dahlman proposes a revision of the semantic notion of presupposition. Presupposition is one of the most debated notions in the linguistic and philosophical literature. Historically, there are two main theoretical approaches to presuppositions. According to the first, the semantic view, presuppositions are semantic implications, that is, truth-conditional relations between propositions and statements. In this sense, presuppositions are considered properties of sentences and a presupposed proposition is a necessary condition for the truth of the presupposing statement. In the second approach, the pragmatic view, presuppositions are not properties of sentences but rather properties of speakers or of linguistic performances given a certain context of utterance. From this view, a presupposed proposition is a condition for the felicitous utterance of the presupposing statement in a given context. Traditionally, it is assumed that semantic presuppositions differ from classical entailments, as presuppositions, unlike classical entailments, project under negation: if we compare a context of entailment to a context of presupposition, we should see that entailments, but not presuppositions, disappear under negation. Following the seminal work of Russell, and against the traditional Strawsonian account of presupposition, Colonna Dahlman argues that presupposition failure does not result in a lack of truth-value, but rather in falsity; hence most standard cases of presuppositions, as discussed in the literature, are classical entailments. Moreover, she claims that all presuppositions that are classical entailments are also pragmatic presuppositions, while not all pragmatic presuppositions are also classical entailments. Colonna Dahlman's claim stands in contrast to Karttunen's well-known analysis of

───────

Territory"), also Capone/Carapezza/Lo Piparo (eds.) (2018), and the monographs and collections in the John Benjamins series *Pragmatics & Beyond*.

factive verbs and his distinction between true factives (that is, emotive factives) and semifactives (that is, cognitive factives).

Andrei Moldovan continues with an investigation of two other classical pragmatic notions: entailment and implicature. He is interested in the relation holding between these two notions, and the question he addresses can be formulated in the following way: whether an entailment *q* of the proposition *p* literally expressed by an utterance of a natural language sentence can, at the same time, figure as the content of a conversational implicature (CI) that the utterance carries. In other words, given that both entailments and conversational implicatures are implications of *what is said*, is it possible for the same proposition to be both entailed and conversationally implicated by what is said? Of course, not all entailments are implicatures, and not all implicatures convey contents that are entailed by what is said. The question is whether this is possible, that is, whether a proposition *could* play this double role. A review of the pragmatic literature that discusses implicatures reveals that the question has received different answers. While those who answer the question affirmatively rely on examples of entailments that look very much like typical cases of CIs, those who answer it negatively adduce considerations pertaining to the conceptual difference that needs to be drawn between the two categories of implications. Moldovan considers the various criteria for identifying CIs, and concludes that both cancellability and non-conventionality seem to exclude entailments from the class of propositions that could be conversationally implicated. However, he also argues that, upon closer examination, neither of the two reasons for a negative answer are compelling. Finally, Moldovan concludes that the answer to the initial question should be affirmative.

In the next chapter, Pedro Santos also discusses conversational implicatures (CIs). He observes that attempts at accounting for so-called embedded conversational implicatures come up against the basic problem of making sense of kinds of content that, despite being derived from *bona fide* conversational implicatures, seem to be part of the literal meaning of utterances. The problem resists the conventionalist way out (championed, among others, by Chierchia and Levinson), which construes generalized CIs as not meriting the status of CIs proper, this being a way of dealing with the puzzle. One decisive drawback of this strategy is that particularized CIs, which could not credibly be conventionalized, display the same kind of puzzling embedding behavior as generalized CIs. Santos discusses Mandy Simons' and François Recanati's approaches to the conundrum. Simons assigns to embedded ICs the status of *bona fide* Gricean inferences, building upon the distinction between embedded pragmatic *effects* and embedded pragmatic *computations* and claiming that the puzzle is solved by the realization that only

the latter phenomenon is really problematic and that embedded ICs are to be an-alysed as cases of the former rather than as cases of the latter. Recanati has qualms about this sort of approach and attempts to account for the phenomenon by dismissing the very notion of an embedded IC as intrinsically incoherent. Re-canati purports to do this on the basis of a globality criterion for Gricean infer-ences which he argues cases of alleged embedded ICs only partially fulfil, so that they count as cases of Gricean inferences *latu sensu* (as, for instance, are cases of what he calls modulation) but surely not as genuine CIs. Santos argues that Re-canati's take on the problem cannot be successful, as globality is a faulty criterion as far as distinguishing cases of modulation from CIs proper is concerned.

In philosophy of language and theoretical linguistics there is an intense de-bate on how utterance meaning is determined. Palle Leth observes that theorists are usually concerned with the semantic component of utterance meaning, i.e. the level of what is said, which cannot be reduced to compositional linguistic meaning, if only because of indexicals. Approaches to this question can broadly be characterized as either intentionalist or anti-intentionalist. Intentionalists take utterance meaning to be determined by the speaker's intention, usually con-strained by conformity to linguistic conventions or by the formation of reasona-ble expectations on being understood. Anti-intentionalists take utterance mean-ing to be determined by public features available to the hearer, such as linguistic conventions and diverse contextual cues. Leth argues that the presupposition that there is such a thing as the meaning of an utterance is unwarranted. The interpretive interaction between a speaker (S) and a hearer (H) does not support the notion of the correct interpretation of an utterance. What matters to commu-nicative success is simply the convergence between S's intention and H's inter-pretation. In cases of non-convergence or divergence, which theorists concoct in order to derive intuitions supporting their favoured account of utterance mean-ing, H basically has two options: H may take an interest in S's intention or H may take an interest in the most reasonable interpretation of the utterance. In the for-mer case, H goes along with whatever meaning S wants her to go along with; in the latter case, H's claim is simply that, irrespective of S's intention, she had the best epistemic reasons to take the utterance the way she did. The former option is H's default option, the latter option is exploited when normative consequences are at stake. Leth stresses that neither of these options has any implications for utterance meaning, and that the notion of correct interpretation is idle in practice and the burden of proof is shifted onto those who posit it as a theoretical entity.

Whereas Leth mentions misunderstanding, confusion, and imagined mean-ing, Chi-Hé Elder devotes her contribution to miscommunication, and especially to negotiating what is said in the face of miscommunication. She observes that in

post-Gricean pragmatics, communication is said to be successful when a hearer recovers a speaker's intended message. On this assumption, proposals for 'what is said' – the semantic, or propositional meaning of a speaker's utterance – are typically centred around the content the speaker aimed to communicate. However, these proposals don't account for the fact that speakers can be deliberately vague, leaving no clear proposition to be recovered, or that a speaker can accept a hearer's misconstrual even though the speaker didn't intend it. In such cases, identifying 'what is said' is more contentious, even though communication is arguably no less successful. Building on recent interactionist approaches to meaning, Elder offers a proposal for 'what is said' in cases of miscommunication, and she provides some rationale for the situation when the speaker is expected to accept a divergent interpretation of his or her initial intention. When we speak, each of our utterances can potentially communicate multiple propositions whose obviousness ranges from strong to weak. The strong ones are the most plausible candidates for 'what is said', but in case they are not recovered, the result of meaning negotiation between interlocutors is expected to align with the speaker's higher-order speech act.

In the next contribution, Brian Ball investigates the relation between playing games and linguistic activity. He tentatively defines games as abstract objects individuated by their rules, thought of as essential to them, and offers an account according to which game-playing situations may be regarded as involving *agents* standing in the *playing* relation to abstract *games* having their rules essentially. Ball inquires whether linguistic activity can be so regarded as well. Timothy Williamson has argued that it can, and that assertion in particular should be thought of as a move in such a game, governed by the rule *assert only what you know*; but Ishani Maitra argues against the possibility of such an analysis. Ball defends the Williamsonian proposal by rejecting accounts of rule-following that would allow the argument to go through. Arguably, playing a game does require that agents follow its rules; but agents do, in the relevant (weaker) sense, follow a strong rule of assertion. To follow a rule one must have some tacit knowledge of it, and Ball suggests that rule-following so construed is involved in the playing of language games. Even if this account of rule-following should prove to be incorrect, it is nevertheless the case that flagrant rule-violations are compatible with game-play.

Joanna Odrowąż-Sypniewska takes a look at another aspect of language games, namely *making moves* through subsentential speech acts. The most commonly given examples of subsentential speech acts are expressions such as "Nice dress", "From Spain", "Where?", etc., uttered in circumstances in which speakers uttering them are regarded as making moves in different language games (such

as stating, asking, promising, etc.). The argument from connectivity is one of the most important arguments for the claim that such utterances – contrary to appearances – are in fact ellipses, i.e. sentential speech acts. The argument uses examples from inflectional languages, such as Polish or German, in which allegedly subsentential speech acts appear in cases other than the nominative. Those who think that they are just fragments of longer unpronounced sentences have no problem in explaining where the case comes from, but the answer is more problematic for those who think that such utterances are truly subsentential. Odrowąż-Sypniewska argues that this argument is by no means conclusive and that defenders of subsentential speech acts need not be worried by connectivity effects. She suggests a situated contextualist account of subsentential speech acts, which is based on Recanati's moderate relativism. Recanati's relativism presupposes two principles: duality and distribution. In the case of subsentential speech acts we have to postulate a two-staged principle of distribution: it's not only the case that the determinants of truth-value distribute over content and circumstance, but also that the content itself is distributed over the locutionary 'what is said' and the situation of the utterance. Odrowąż-Sypniewska argues – *pace* Recanati and following Perry – that, at least in the case of subsentential speech acts, we have to postulate unarticulated constituents in explicit contents as well as in the situations of evaluation. The explanation of connectivity effects appeals to Perry's idea of the utterance completing the thought whose other constituents are already in the situation. The speakers use cases other than the nominative in order to simplify the process of enrichment for the hearers. The cases make it easier to determine which completion of the articulated content is the intended one. Odrowąż-Sypniewska concludes that the contextualist account is able to withstand the objection from connectivity and can be regarded as a view that explains how it is that subsentential expressions can be used to make a move in a language game.

Maciej Witek concentrates on the process of accommodation in linguistic interaction. Accommodation is a process whereby the context of an utterance is adjusted or repaired in order to maintain the default assumption that the utterance constitutes an appropriate conversational move of a certain type. It involves a kind of redressive action on the part of the audience and, depending on what the appropriateness of a speech act requires, results in providing missing contextual elements such as referents for anaphoric expressions, presuppositions, suppositions, deontic facts, pragmatically enriched contents, etc. It remains to be determined, however, what is the source of the contextual requirements whose recognition motivates and guides the accommodating context-change. Witek addresses this particular question – which expresses the so-called *triggering* or *constitution*

problem – and suggests that it can be adequately answered by a speech act-based model, the central idea of which is that the requirements in question are structural components of patterns, scripts, or procedures for the performance of speech acts. He also argues that the proposed framework can be used to explain a wide range of accommodating phenomena and can shed some new light on the constitution of accommodation-triggering requirements.

Leopold Hess devotes his paper to expressive meanings and expressive commitments. Expressives, i.e. words such as *damn* or *bastard*, seem to convey a specific kind of content, different from, or on top of, "regular" descriptive meaning. Following the seminal work of Chris Potts the meaning of expressives is often conceptualized in a two-dimensional semantic framework, in which descriptive and expressive contents are separated as a result of special rules of semantic composition. This approach is successful in accounting for some interesting semantic properties of expressives and has also been extended to other classes of expressions, such as racial slurs or honorifics. However, it does not offer any actual insight into the nature of expressive meaning (the two-dimensional formalism operates on dummy values, independently of what they may stand in for). Hess offers an alternative, pragmatic, account of expressives, based on the observations that 1) expressive meanings seem directly to involve the speakers (their states, emotions or attitudes) rather than just abstract (e.g. truth-conditional) contents, and that 2) the utterers of expressives are responsible for the choice of loaded, often taboo vocabulary. The presented account is developed in a commitment-based scorekeeping model of discourse (inspired by Lewis and Brandom), in which hearers interpret speakers' utterances by attributing commitments to them. Besides assertoric commitments (and potentially other kinds), expressive commitments can be distinguished. These are commitments to the appropriateness or applicability of a given expression, which also can be attributed to speakers based on their utterances (separately from assertoric commitments).

Pedro Abreu aims to assess Donald Davidson's arguments against the sufficiency and necessity of conventions for successful linguistic communication. Davidson goes beyond the common contention that the basic conventional layer of meaning, one that is secured by interlocutors' shared competence in their common language, must often be supplemented in rich and inventive ways. First, he maintains that linguistic understanding is never exclusively a matter of mere decoding, but *always* an interpretative task that demands constant additional attention to the indeterminately various cues and clues available. More radically still, Davidson denies that linguistic conventions are even needed. In particular, he argues against the fairly consensual thesis there is some essential element of

conventionality in literal meaning. This still represents a very distinctive contribution to the persistent and tumultuous discussion over the relative natures and limits of semantics and pragmatics. Abreu maintains that Davidson is only partially right in his claims and develops an argument supporting the thesis that genuine pursuit of linguistic understanding can never take the form of uncritical conformity to a fixed norm. Finally, Abreu considers in some detail Davidson's argument from radical interpretation and concludes that it fails.

In the final contribution, Tadeusz Ciecierski tackles one of the classical problems of philosophy of language (and philosophy in general): the relation between knowledge and belief. Ciecierski defends the orthodox view that knowledge entails belief. He distinguishes two ways in which one can deny the knowledge-belief entailment claim (compatibilism and incompatibilism). Compatibilism is the claim that the co-occurrence of knowledge and belief is a purely contingent matter, while incompatibilism states that co-occurrence is impossible. Next, Ciecierski presents intuitive arguments against both kinds of denials, and discusses some possible objections against the presented arguments. Finally, he concludes that there are reasons to think that the traditional picture is correct; it seems, therefore, that we should remain sceptical about the philosophical prospects of compatibilism and incompatibilism.

The chapters in this volume fully justify the title of the collection – *Philosophical Insights into Pragmatics*. It would also be possible to reverse the wording, and envisage a volume on *pragmatic insights into philosophy*, which remains a viable project for the future.

Acknowledgments

References

Capone, Alessandro/Carapezza, Marco/Lo Piparo, Franco (Eds.) (2018): *Further Advances in Pragmatics and Philosophy. Part 1 From Theory to Practice* (Perspectives in Pragmatics, Philosophy & Psychology 18), Cham: Springer.

Capone, Alessandro/Lo Piparo, Franco/ Carapezza, Marco (2013): "Introduction". In: Alessandro Capone/Franco Lo Piparo/Marco Carapezza (Eds.): *Perspectives on Pragmatics and Philosophy*, Cham: Springer, pp. xi–xvii.

Capone, Alessandro/Lo Piparo, Franco/Carapezza, Marco (Eds.) (2013): *Perspectives on Pragmatics and Philosophy*. (Perspectives in Pragmatics, Philosophy & Psychology 1), Cham: Springer.

Caton, Charles (1971): "Overview". In: Danny Steinberg/Leon Jakobovits (Eds.): *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge: Cambridge University Press, pp. 3–13.

Depraetere, Ilse/Salkie, Raphael (Eds.) (2017): *Semantics and Pragmatics: Drawing a Line* (Logic, Argumentation & Reasoning 11). Cham: Springer.

Jaszczolt, Kasia M. (2018): "Pragmatics and Philosophy: In Search of a Paradigm." In: *Intercultural Pragmatics* 15(2), pp. 131–159. DOI: https://doi.org/10.1515/ip-2018-0002

Kempson, Ruth (1977): *Semantic Theory* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Sbisà, Marina/ Östman, Jan-Ola/Jef Verschueren (Eds.) (2011): *Philosophical Perspectives for Pragmatics* (Handbook for Pragmatics Highlights 10). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Stalmaszczyk, Piotr (Ed.) (2014): *Semantics and Beyond. Philosophical and Linguistic Inquiries* (Philosophical Analysis 57). Berlin/Boston: De Gruyter.

Turner, Ken (2011): "Introduction: Preliminary 'sketches of landscape'". In: Ken Turner (Ed.): *Making Semantics Pragmatic* (Current Research in the Semantics/Pragmatics Interface. Vol. 24). Bingley: Emerald Group Publishing Limited, pp.1–18.

Turner, Ken (Ed.) (2011): *Making Semantics Pragmatic* (Current Research in the Semantics/Pragmatics Interface. Vol. 24). Bingley: Emerald Group Publishing Limited.

Verschueren, Jef/ Östman, Jan-Ola (Eds.) (2011): *Key Notions for Pragmatics* (Handbook for Pragmatics Highlights 1). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Woleński, Jan (2003): *Epistemologia. Tom III. Prawda i realizm* (Epistemology. Vol. 3. Truth and realism). Kraków: Aureus.

Roberta Colonna Dahlman

Presuppositions, again

Abstract: *Presupposition* is surely one of the most debated notions in the linguistic and philosophical literature. Historically, there are two main theoretical approaches to presuppositions. According to the first one, the semantic view, presuppositions are semantic implications, that is, truth-conditional relations between propositions and statements. In this sense, presuppositions are considered properties of sentences and a presupposed proposition is a necessary condition for the truth of the presupposing statement. In the second approach, the pragmatic view, presuppositions are not properties of sentences but rather properties of speakers or of linguistic performances given a certain context of utterance. From this view, a presupposed proposition is a condition for the felicitous utterance of the presupposing statement in a given context.

Traditionally, it is assumed that semantic presuppositions differ from classical entailments, as presuppositions, unlike classical entailments, project under negation: if we compare a context of entailment to a context of presupposition, we should see that entailments, but not presuppositions, disappear under negation. This presentation aims to propose a revision of the semantic notion of presupposition. I argue that most standard cases of presuppositions are classical entailments. Moreover, I claim that all presuppositions that are classical entailments are also pragmatic presuppositions, while not all pragmatic presuppositions are also classical entailments. I contend that factive verbs offer a paradigmatic example of this distinction, as the factivity related to *know* is semantic, hence a classical entailment, whereas the factivity related to *regret* is merely pragmatic. This claim stands in contrast to Karttunen's (1971) well-known analysis of factive verbs and his distinction between *true factives* (that is, emotive factives) and *semifactives* (that is, cognitive factives).

Keywords: presupposition, classical entailment, constancy under negation, projection, factivity, *know*, *regret*

1 Introduction

This paper aims to propose a revision of the semantic notion of presupposition. Following Russell (1905), and against the traditional Strawsonian account of presupposition, I argue that presupposition failure does not result in a lack of truth-

value, but rather in falsity, and hence that most standard cases of presuppositions, as discussed in the literature, are classical entailments. Moreover, I claim that all presuppositions that are classical entailments are also pragmatic presuppositions, while not all pragmatic presuppositions are also classical entailments. I contend that factive verbs offer a paradigmatic example of this distinction, as the factivity related to *know* is semantic, that is, a classical entailment, whereas the factivity related to *regret* is a merely pragmatic phenomenon. This claim stands in contrast to Karttunen's (1971) well-known analysis of factive verbs and his distinction between *true factives* and *semifactives*.

2 Presupposition

Presupposition is surely one of the most debated notions in the linguistic and philosophical literature.[1] Typical contexts of presupposition are the following:

Definite descriptions:
(1) The King of France is bald.
     presupposes: 'There is a (unique)[2] King of France.'
Clefts:
(2) It was John who wrote a letter to Mary.
     presupposes: 'Someone wrote a letter to Mary.'
Change of state verbs:
(3) John stopped running (at time t_1).
     presupposes: '(Before t_1) John was running.'
Factives:
(4) John knows that Mary lives in Sweden.
     presupposes: 'Mary lives in Sweden.'

It is commonly assumed that presuppositions are not classical entailments, as presuppositions, unlike classical entailments, survive negation (*constancy under negation*). If we compare a context of entailment to a context of presupposition, we should see that entailments, but not presuppositions, disappear under negation:

---

1 For an overview of different theories on presupposition, see Dinsmore (1981), Levinson (1983), Soames (1989), Huang (2007), Beaver and Geurts (2013), Simons (2013).
2 Note that, in this study, I am not considering the implication of uniqueness.

(5) a. Beatrix killed Bill.
      entails: 'Bill is dead.'
      b. Beatrix did not kill Bill.
      does not entail: 'Bill is dead.'
(6) a. The King of France is bald.
      presupposes: 'There is a King of France.'
      b. The King of France is not bald.
      presupposes: 'There is a King of France.'

The sentence in (5a) entails the proposition 'Bill is dead', whereas its negative counterpart in (5b) does not. On the other hand, it is assumed that the sentence in (6a) implies an existential presupposition ('There is a King of France'), and that its negative counterpart in (6b) also does. *Projection* is the term used to denote this typical property of presuppositions, that is, the capacity to survive under negation.[3]

Moreover, it is assumed that presuppositions project under other entailment-cancelling operators, namely questions, conditionals and modality operators, as shown in (7)–(9):

(7) a. Is the King of France bald?
      presupposes: 'There is a King of France.'
      b. Did Beatrix kill Bill?
      does not entail: 'Bill is dead.'
(8) a. If the King of France is bald, we will need to buy him a hat.
      presupposes: 'There is a King of France.'
      b. If Beatrix killed Bill, we will need to make arrangements for his funeral.
      does not entail: 'Bill is dead.'
(9) a. Maybe the King of France is bald.
      presupposes: 'There is a King of France.'
      b. Maybe Beatrix killed Bill.
      does not entail: 'Bill is dead.'

---

3 The "projection problem" for presuppositions was originally formulated by Langendoen and Savin (1971) as the question of "how the presupposition and assertion of a complex sentence are related to the presuppositions and assertions of the clauses it contains" (p. 55). The projection principle was stated as follows: "presuppositions of a subordinate clause [...] stand as presuppositions of the complex sentence in which they occur" (Langendoen & Savin 1971, p. 57). In other words, according to this principle, a complement clause should inherit all the presuppositions of the complex sentence in which it occurs (so-called *cumulative hypothesis*).

Chierchia and McDonnell-Ginet (1990, p. 24) argue that whenever S presupposes P, there is an entire family of sentences related to S (so-called *S family*), namely its negative variant, its interrogative variant, its epistemic modal variant, and the conditional with S as its antecedent, which also *tend* to presuppose P.[4]

Historically, there are two main theoretical approaches to presuppositions. According to the first one, the semantic view, presuppositions are semantic implications, that is, truth-conditional relations between propositions and statements. In this sense, presuppositions are considered properties of sentences or statements (that is, uses of sentences).[5] In the second approach, the pragmatic view, presuppositions are not properties of sentences but rather properties of speakers or of linguistic performances given a certain context of utterance. From the semantic view, a presupposed proposition is a necessary condition for the truth of the presupposing statement; while from the pragmatic view, a presupposed proposition is a condition for the felicitous utterance of the presupposing statement in a given context, a "felicity condition of statemenhood", as Ken Turner (1992, p. 369) puts it.

The currently dominant notion of presupposition is a pragmatic notion. In the following sections, I show why the pragmatic view turned out to be the most convincing account, and then I put forward arguments in favour of a semantic notion in terms of classical entailment. I argue that most standard cases of presupposition, besides being pragmatic implications, are classical entailments, and that to overlook this hypothesis results in a disregard of crucial distinctions.

---

4 Crucially, we will see that presupposition projection is not a logical necessity, but depends on the context of utterance. In most cases, the context of utterance is such that presuppositions project in embedded entailment-cancelling environments. However, it can be the case that projection does not prevail. In his well-known study on *holes*, *plugs* and *filters*, Karttunen (1973) proved that the projection principle is too strong: in his line of reasoning, some presuppositions project (because they are embedded under *holes*), while for others the mechanism of projection is necessarily blocked (because they are embedded under *plugs*) or is contextually – "under certain conditions" – blocked (because they are embedded under *filters*). In most recent works (see references in section 9), it has been argued that projection tests address information structure level and are a diagnostic for the discourse status of implications rather than for presuppositionality.

5 On the distinction between statements and sentences, see Strawson (1952, p. 174 ff.). See also Hintikka (1962, p. 6): "[..] a statement is the act of uttering, writing, or otherwise expressing a declarative sentence. A sentence is the form of words which is uttered or written when a statement is made."

3 The semantic view

The notion of presupposition was introduced into linguistic theory and philosophy of language by the logician and philosopher Gottlob Frege (1892, Eng. translation 1948). In *Über Sinn und Bedeutung* (1892, Eng. translation 1948), Frege wrote:

> [..] when we say "the moon", we do not intend to speak of our conception of the moon, nor are we satisfied with the sense alone, but we *presuppose* a referent.
>
> <div align="right">(Frege 1948, p. 214, emphasis mine)</div>

And further on:

> If anything is asserted there is always an obvious presupposition that the simple or compound proper names used have referents. If one therefore asserts "Kepler died in misery," there is a presupposition that the name "Kepler" designates something; but it does not follow that the sense of the sentence "Kepler died in misery" contains the thought that the name "Kepler" designates something. [..] That the name "Kepler" designates something is just as much a presupposition for the assertion Kepler died in misery as for the contrary assertion.
>
> <div align="right">(Frege 1948, pp. 221–222)</div>

In a similar vein, Peter F. Strawson (1950, 1952, p. 175 ff.) defined presupposition as a relation of *truth dependence* between statements, that is, uses of sentences. This view shifted to the definition of *semantic presupposition* as the relation between statements that can be true or false, hence can express *propositions*: if the presupposing proposition is true, then the presupposed proposition *must* also *be true* (i); if the presupposed proposition is false, then the presupposing proposition is *neither true nor false* (ii).

(10) The Strawsonian notion of semantic presupposition
      S presupposes P iff:
      (i) if S then it *must* be the case that P (= P is a truth condition of S);
      (ii) if ¬P then S has *no truth-value* (that is, S is neither true nor false).

According to this notion, if S presupposes P, then in case of presupposition failure (¬P), S is *neither true nor false*, meaning that S cannot be used to make an assertion that is either true or false. As Strawson writes, the question whether S is true or false does not arise (Strawson 1950, p. 330, see also Geach 1950, p. 85). This is the point on which Strawson famously criticized Russell's theory of denoting phrases. According to Bertrand Russell (1905), a sentence such as "The present King of France is bald" is simply false, as there is no King of France; while,

in Strawson's line of reasoning, the same sentence has no truth-value – it is neither true nor false.[6]

As a next step, Bas van Fraassen (1968, p. 137), following the Frege-Strawsonian view, defined presupposition in purely semantic terms:

> The explicit characterization of *presupposes* is therefore given by
> 1. *A presupposes B* if and only if *A* is neither true nor false unless *B* is true.
> This is equivalent to
> 2. *A* presupposes *B* if and only if
> (a) if *A* is true then *B* is true,
> (b) if *A* is false then *B* is true.

Along the same line, Keenan (1971, p. 45) defined a *logical presupposition* as follows:

> A sentence S logically presupposes a sentence S' just in case S logically implies S' and the negation of S, ~S, also logically implies S'. In other words, the truth of S' is a necessary condition on the truth or falsity of S.

Eventually, the semantic notion of presupposition was represented in the following terms:

(11) The traditional semantic notion of presupposition
      S presupposes P iff:
      (i) if S then P (= P is a truth condition of S);
      (ii) if ¬S then P (= P is a truth condition of ¬S).

The assumption of *constancy under negation* stems from this semantic notion as grounded in the Frege-Strawsonian account of presupposition. The claim was

---

6 On the truth-value gap theory, see also Quine (1960), Austin (1962). Later, Strawson (1964) claimed that whether S is false or truth-value lacking is a matter of *topicality*: if presupposition failure affects the topic of the statement (what the statement is about, as in "The King of France is bald"), then S is neither true nor false; while, otherwise, if presupposition failure does not affect the topic of the statement, but merely affects what purports to be information about its topic (as in "The exhibition was visited by the King of France", which answers the question "Who visited the exhibition?"), then S can be said to be false, as the statement is misinformative about its topic (see also Strawson 1954, p. 226). In this sense, see also Reinhart (1981, pp. 69–70), Erteschik-Shir (1997, p. 16). For an alternative view, see von Fintel (2004), who claims that "All the sentences referring to the present king of France should equally fail to assign a truth-value to our world" (p. 326). In defense of Strawson's discourse-based account of presupposition failure, and against von Fintel's analysis, see, more recently, Bezuidenhout (2016).

that in order to assign a truth-value (true/false) to S, P must be true. Therefore, the conclusion that S *entails* P, and ¬S *entails* P. Both the statement "The King of France is bald" and its negative counterpart "The King of France is not bald" *must* imply that "There exists a King of France." In this sense, presuppositions were assumed necessarily to project under negation.

In this light, *Modus Tollens* was not assumed to apply to presuppositions, as shown by condition (iii) in (12):

(12) The traditional semantic notion of presupposition (II)
    S presupposes P iff:
    (i) if S then P (= P is a truth condition of S);
    (ii) if ¬S then P (= P is a truth condition of ¬S);
    (iii) if ¬P then ●S (that is, S is neither true nor false).[7]

That *Modus Tollens* does not apply to presuppositions means that presuppositions must be considered as being clearly distinguished from classical entailments.

4 The pragmatic view

The failure of the traditional semantic notion of presupposition, where both S and ¬S necessarily imply P, was largely demonstrated by several scholars who showed how presuppositions may vanish under negation and other entailment-cancelling operators, such as questions, conditionals and possibility operators, rather than being constant.[8] Therefore, since the 1970s, presuppositions have received a pragmatic account in linguistic theory and have been treated as phenomena mainly related to the felicity conditions of the utterance (Karttunen 1970, 1971, 1973, 1974) or to the speaker's subjectivity (Stalnaker 1972, 1973, 1974, 1999,

---

7 The symbol ● stands for "neither true nor false".
8 See, for instance, Karttunen (1971), who pointed out how some factive presuppositions may vanish, instead of being constant, in some environments, despite the assumption that "Whatever a sentence with a factive predicate presupposes, *the presupposition ought to remain* no matter whether the main sentence is a negative assertion, an interrogative sentence, or the antecedent of a conditional construction" (Karttunen 1971, pp. 62–63, emphasis mine). Note that it is quite remarkable that these scholars, while criticizing the semantic notion of presupposition, seemed to accept the premise established by this very notion, namely that *presuppositions should be constant under negation*.

2002), and not to the truth-conditional content of the uttered sentence. The traditional semantic notion of presupposition, founded on the idea that presuppositions are truth conditions for statements, was replaced by a pragmatic notion grounded on the idea that presuppositions are conditions on the context of utterance – they must be met in the context of utterance in order for a sentence to be appropriately uttered.

The pragmatic notion of presupposition currently dominant in the linguistic literature is credited to Robert Stalnaker (1972, 1973, 1974, 1999, 2002). According to Stalnaker, presupposition should be defined as a relation *between speakers and propositions*. As Stalnaker puts it, it is speakers who presuppose, not sentences or propositions:

> The notion I will discuss is a *pragmatic* notion, as opposed to a purely *semantic* one. This means that the presupposition relation cannot be explained solely in terms of the meaning or content of sentences, but must be explained partly in terms of facts about users of sentences: their beliefs, intentions and expectations. [...] [T]he basic presupposition relation is not between propositions or sentences, but between a person and a proposition.
>
> (Stalnaker 1973, p. 447)

Regarding the Stalnakerian view, a crucial point needs to be clarified. According to this view, the term 'presupposition' is used ambiguously to denote two different things. On the one hand, Stalnaker defines presupposition as a *propositional attitude, a mental state of the speaker*: "According to the *pragmatic* conception, presupposition is a propositional attitude, not a semantic relation" (Stalnaker 1972, p. 387). Presuppositions, as Stalnaker says, constitute the background of knowledge or beliefs (the so-called *common ground*) against which statements and requests are made, questions are asked, and proclamations and commands are issued (Stalnaker 1973, p. 448). Moreover, he explicitly underlines that presupposing is not a mental attitude like believing, but rather "a *linguistic disposition* – a disposition to behave in one's use of language *as if* one had certain beliefs, or were making certain assumptions" (Stalnaker 1974: 202, emphasis mine, 1999):[9]

---

9 Stalnaker defines this disposition in terms of *acceptance*: "To accept a proposition is to treat it as true for some reason" (Stalnaker 2002, p. 716). The Stalnakerian notion of acceptance was introduced in Stalnaker (1973): "[...] it is clear that presupposition is a propositional attitude. More specifically, it is an attitude of accepting something to be true" (p. 450). For more on this notion, see also Stalnaker (1984).

(13) The Stalnakerian notion of pragmatic presupposition as relation between speakers and propositions
S presupposes P iff:
The speaker of S, in uttering S, assumes or acts *as if* she assumes that P is true and that the truth of P is part of the *common ground* of information shared with her interlocutor(s).

According to this view, as I understand it, presupposition is the speaker's act of taking the truth of a proposition as part of the conversational common ground at the time of her utterance.[10] This act may or may not be justified by the speaker's own belief depending on whether the speaker herself believes that the presupposed proposition be entailed by the common ground or not – in this latter case, she is just acting *as if* the presupposed proposition were part of the common ground.

On the other hand, Stalnaker uses the term presupposition to indicate *the proposition whose truth is taken for granted*: "A person's presuppositions are the propositions whose truth he takes for granted, often unconsciously, in a conversation, an inquiry, or a deliberation" (Stalnaker 1973, p. 447). In this sense, presuppositions are those *propositions* that constitute the *common ground* of information shared by speakers and hearers, they are *the background of information against which assertions and other speech acts are made*:

(14) The Stalnakerian notion of pragmatic presupposition as backgrounded information
S presupposes P iff:
In uttering S, P is not asserted but backgrounded information, as it already belongs to the common ground shared by speakers and hearers.

This definition is based on a distinction made on the level of *information structure*: the distinction between presupposition as backgrounded information and assertion as foregrounded information (see Hooper 1975). In this sense, what is asserted in an utterance is the answer to some *question under discussion* (cf. Simons *et al.* 2010), while the presupposed content of an utterance is not at stake or open for discussion, but rather taken for granted.

One of the advantages offered by the pragmatic notion of presupposition is that this framework allows us to distinguish between the level of sentence meaning and

---

10 For arguments against Stalnaker's *common ground* theory of presupposition, see Abbott (2000, 2008, 2016), Simons (2001, 2005, 2006), and Gauker (2008, 2016).

the utterance level. In this view, it is possible to see that some presuppositions, in fact, most standard cases of presupposition, are also entailments (for this sense, see Stalnaker 1974, 1999, Chierchia & McConnell-Ginet 1990, Heim 1992, Seuren 2006, Abrusán 2011, Abbott 2016). As Stalnaker (1999, p. 54) puts it: "[…] if presupposition is defined independently of truth-conditions, then one can separate the question of entailment relations from the question of presupposition". This possibility was excluded by the traditional semantic view, because presuppositions, unlike entailments, were considered necessarily constant under negation. According to the semantic notion, if S presupposes P, then both S and ¬S entail P, meaning that P is a truth-condition both of S and ¬S. According to the pragmatic notion, if a speaker, by uttering S, presupposes P, then it is the case (i) that S *entails* P (which means that ¬S does not entail P), and it is possible (ii) that ¬S *presupposes* P. In this framework, presuppositions, unlike entailments, *tend* to project in embedded environments (that is, sentences embedded under negation, questions, conditionals, and modality operators), but they do not always project, as sometimes they are *cancelled*.[11] In this view, it is assumed that the relation between S and P and the relation between ¬S and P are different kinds of implications, an entailment in the former case, and a cancellable pragmatic implication in the latter. As Marina Sbisà (1999, p. 331) puts it,

> The relationship of the presupposition to the utterance of the positive sentence appears not to be identical to its relationship to the utterance of the negative sentence, since in the latter case the presupposition is cancelable.

Most recent work on presuppositions aims at drawing a line of distinction between presuppositions that are also conventionally entailed content and presuppositions that are not entailed content (for an overview of these current theories, see Djärv *et al.* 2016). All these accounts, however, seem to agree on the view that presuppositions cannot be classical entailments. This view was recently expressed by Barbara Abbott, who writes (Abbott 2016, p. 10):

> The presupposed propositions in (1) and (2) are also entailed by their presupposing sentences; if the king of France is wise is true, then there must be a king of France.

---

11 Being merely pragmatic implications, presuppositions are assumed to be cancellable. On the notion of potential (or putative) presupposition and the so-called "cancellation analysis" of presuppositions, see Gazdar (1979b, p. 64 ff.), Huang (2007, p. 81), Levinson (1983, p. 186 ff.), Soames (1989, p. 573 ff.). Other terms have been used to denote the same phenomenon of presupposition cancellation/cancellability: for instance, Abrusán (2016) and Macagno and Capone (2016) use the term "suspension", and Abbott (2016) uses the term "neutralizability".

Then, in footnote 1 (Abbott 2016, p. 10, emphasis mine), she clarifies that

> If presupposition failure results in lack of a truth value for the presupposing sentence (as believed by many people, in addition to Frege and Strawson), then the relation between that sentence and its presuppositions *can't be classical entailment*, since modus tollens would no longer hold in this case. I will continue to use the terms "entail" and "entailment", assuming this proviso.

The aim of this paper is to argue against this view, and to propose that most standard cases of presupposition, if considered on a semantic level of analysis, *are* classical entailments.[12]

5 What was wrong with the traditional semantic view

We have seen that the failure of the semantic notion of presupposition, where both S and ¬S necessarily imply P, was largely demonstrated by several scholars, who showed how presuppositions may vanish under negation and other entailment-cancelling operators, such as questions, conditionals, and possibility operators, rather than being constant. This behaviour could have been seen as evidence against the alleged distinction between presuppositions and entailments, but instead the option of a pragmatic notion of presupposition was preferred. The line of reasoning was the following:

(15) The argument against the semantic view
(i) According to the semantic notion, presuppositions are implications that are clearly distinguished from classical entailments, as the former, but not the latter, are constant under negation (and other entailment-cancelling operators).
(ii) It is not the case that presuppositions are always constant under negation (and other entailment-cancelling operators).
(iii) Therefore, the semantic notion of presupposition does not hold, and a pragmatic notion is preferable.

---

12 In this sense, see also the analysis presented by Stanosz (1991, p. 98 ff., Swedish translation from Polish by Semantix Språkcentrum AB). I am indebted to Tadeusz Ciecierski, who brought Barbara Stanosz's work to my attention.

Against this line of reasoning, I propose the following alternative:

(16)   The argument for a revisited semantic view
   (i) According to the traditional semantic notion, presuppositions are impli-
   cations that are clearly distinguished from classical entailments, as the for-
   mer, but not the latter, are constant under negation (and other entailment-
   cancelling operators).
   (ii) It is not the case that presuppositions are always constant under nega-
   tion.
   (iii) Therefore, presuppositions are not implications that are clearly distin-
   guished from classical entailments.

My claim is that most standard cases of presupposition *are* classical entailments.
I argue that the problem with the traditional semantic analysis of presuppositions
is that it resulted in confusing the semantic level of truth-conditional meaning
and the pragmatic level of assertibility (truth conditions vs. conditions of assert-
ibility; cf. Sellars 1954, p. 202 ff.; Neale 1990, pp. 27–28).[13] To say that S is neither
true nor false in case of presupposition failure "as the question whether S does
not arise at all" is not a claim about S's truth or falsity, but rather a claim about
our own incapacity to evaluate S's truth or falsity.[14] It is a claim about the feeling
of "squeamishness" we may perceive if forced to say whether S is true or false
(Strawson 1950, p. 330; von Fintel 2004, p. 321). It is a claim that does not state
the lack of a truth-value, but rather the *lack of interest in raising the question
whether S in a conversation*. To say that S is neither true nor false in case of pre-
supposition failure is not a claim about the physical reality described by S, but
rather a claim that describes a *normative* reality, given that S cannot be evaluated
by the actors of a communicative situation: it is to say that *a speaker is not at
liberty to use S*, hence that if a speaker uses S her communicative act is "out of
order" (cf. Sbisà 1999, p. 331). That a speaker is not at liberty to use S says nothing

---

13 In his review article on Burton-Roberts' revised theory of semantic presupposition (N. Bur-
ton-Roberts, *The Limits to Debate: A Revised Theory of Semantic Presupposition*. Cambridge:
Cambridge University Press, 1989), Turner (1992, p. 368 ff.) stresses the importance of this dis-
tinction and suggests that both Strawson and Burton-Roberts actually defend a *pragmatic* con-
ception of presupposition.
14 Here, I am assuming that a clear distinction has to be drawn between *knowing the truth con-
ditions of a sentence* (that is, knowing what must be the case in the world in order for this sen-
tence to be true or false) and *evaluating the truth-value of a sentence* (that is, assessing whether
the sentence is true or false). On this distinction, see Casalegno (1998, p. 12).

about whether S is true or false, that is, whether S is a correct description of the physical reality or not.

By distinguishing between truth conditions and conditions of assertibility, I argue that presupposition failure is a source of *inappropriateness* for the utterance of the presupposing sentence, which is a claim on a normative reality, and of *falsity* of the presupposing sentence, which is a claim concerning physical reality. In other words, I propose the following semantic account of presupposition:

(17) The semantic notion of presupposition, revisited
    S presupposes P iff:
    (i) if S then P (= P is a truth condition of S);
    (ii) if ¬P then ¬S.

In this view, *Modus Tollens* applies to presuppositions, thus from a semantic point of view there is no reason to distinguish presuppositions from classical entailments. Furthermore, I assume that presupposition failure (¬P) leads to utterance inappropriateness; however, I do not consider this condition as a matter of truth-conditional semantics.

The line of reasoning developed by Strawson, and then by van Fraassen and Keenan, overlooked the above-mentioned distinction between truth conditions and conditions of assertibility. As we have seen, according to these scholars, the claim that S presupposes P if and only if S is neither true nor false unless P is true is equivalent to the claim that if S then P, and if ¬S then P. The problem with this view consists in having assumed an equivalence holding between the claim that *S is neither true nor false unless P is true* and the claim that *S is either true or false if P is true*. This assumption does not hold. *Neither-true-nor-false S* is not equivalent to the negation of true S (that is, ¬S) and the negation of false S (that is, S), but rather to *not evaluable (and inappropriate) S*. The semantic notion of presupposition, in its traditional formulation, was grounded on this equivocation: given the claims that

(a) S is neither true nor false unless P is true, and that
(b) if S is true then P is true, and if S is false then P is true,

the term 'neither true nor false' in (a) does not denote a (lack of a) truth-value, but rather a (negative) condition of assertibility, while the terms 'true' and 'false' in (b) denote truth-values.

6 My proposal (following Russell 1905)

In the 1950s, Strawson famously criticized Russell's theory of denoting phrases. According to Russell (1905), a sentence such as "The present King of France is bald" is simply false, as there is no King of France; in Strawson's line of reasoning, the same sentence has no truth-value – it is neither true nor false. In this paper, I call into question the Strawsonian account and follow Russell in claiming that presupposing sentences are false in case of presupposition failure.

This analysis seems to have problems in explaining the fact that negative presupposing sentences sometimes entail their presupposition, as formulated in (18):

(18) The semantic notion of presupposition, revisited (II)
S presupposes P iff:
(i) if S then P;
(ii) if ¬P then ¬S;
(iii) if ¬S then P or ¬P (that is, ¬S may entail P).

Indeed, given the account as stated in (18), an objection may be formulated as follows:[15]

(19) Objection to the revisited semantic notion of presupposition
Given that:
S semantically presupposes P iff:
(i) S entails P;
(ii) ¬S may entail P;
(iii) ¬P entails ¬S.
Suppose P is false. It follows that:
(a) By (i) and (iii), *S is false* (¬S is true).
(b) By (ii) and (iii), *¬S is false* (S is true).
It is contradictory to say that both S and ¬S are false.

My counterargument to this objection is the following:

---

15 I am indebted to Christopher Gauker, who directed my attention to this objection.

(20) Reply to the objection in (19)

There is no contradiction because ¬S in (a) stands for the *wide scope negation of S*, while ¬S in (b) stands for the *narrow scope negation of S*. In other words, ¬S in (a) and ¬S in (b) refer to two different logical forms.

Following Russell's Theory of Descriptions (1905), and given the negative sentences in (21)–(24), my proposal is to analyse them as semantically ambiguous between a reading where the negation operator has a narrow scope (a-sentences in (21)–(24)) and a reading where the negation operator has a wide scope (b-sentences in (21)–(24)):

(21) The King of France is not bald.
a. There is a King of France such that he is not bald.
b. There is no King of France such that he is bald.

(22) It was not John who wrote a letter to Mary.
a. There was an act of writing a letter to Mary such that John did not do it.
b. There was no act of writing a letter to Mary such that John did it.

(23) John did not stop running.
a. There was an act of running by John such that John did not stop it.
b. There was no act of running by John such that John stopped it.

(24) John does not know that Mary lives in Sweden.
a. There is a true proposition p (that Mary lives in Sweden) such that John does not know it.
b. There is no true proposition *p* (that Mary lives in Sweden) such that John knows it.[16]

As we can see in (21)–(24), the reading of ¬S where the negation operator has a narrow scope entails P, while the reading of ¬S where the negation operator has a wide scope does not.[17] Moreover, we can see that the reading of ¬S where the

---

[16] The analysis in (24) may strike us as problematic. The problem is that the wide scope negation reading in (24b) is meant to cover examples such as (i), but also examples such as (ii):
(i) John does not know that Mary lives in Sweden. In fact, Mary lives in Italy.
(ii) John does not know that Mary lives in Sweden. He has no idea on Mary's whereabouts, and neither do I.

This is quite interesting and surely worth of further investigation. For now I suggest that the wide scope negation reading in (24b) is *unspecified* between a reading where the speaker is committed to the truth of ¬*p*, and a reading where the speaker is not committed to the truth of *p*.

[17] In Russell's terms, we could say that the sentences in a. contain *primary occurrences*, while those in b. contain *secondary occurrences* of denoting phrases (an entity which is now King of

negation operator has a wide scope entails the falsity of the reading of ¬S where the negation operator has a narrow scope. In other words, if it false that there is a King of France, then it is false that the King of France is bald, in the sense that 'There is no King of France such that he is bald', which also entails the falsity of 'There is a King of France such that he is not bald'; if it is false that Mary lives in Sweden, then it is false that John knows that Mary lives in Sweden, in the sense that 'There is no true proposition p[Mary lives in Sweden] such that John knows it', which also entails the falsity of 'There is a true proposition p[Mary lives in Sweden] such that John does not know it.'[18] Thus, the revisited semantic notion of presupposition in terms of classical entailment is illustrated as follows:

(25) The semantic notion of presupposition, revisited (III)
S presupposes P iff:
(i) if S then P;
(ii) if ¬S$_{(WIDE)}$ then ¬P;
(iii) if ¬S$_{(NARROW)}$ then P;
(iv) if ¬P then ¬S$_{(WIDE)}$.

Of course, to claim that most standard cases of presupposition are classical entailments means that the conditions represented in (25) must hold for standard cases of classical entailment, too. In other words, we must expect that standard entailment contexts, when under negation, are scopally ambiguous, and that the entailment survives if under the scope of narrow negation. I contend that this is the case, as shown in (26):

France, a past act of writing a letter to Mary, a past act of running performed by John, a true proposition p).

18 An anonymous reviewer observes that my account seems to say that "falsity of presupposition entails the non-truth (wide-scope negation) of the falsity of S, which [..] appears to be entirely compatible with lack of truth value of S, thereby smuggling in the Strawsonian analysis through a back door." Indeed, my account would be compatible with the Strawsonian analysis postulating lack of truth-value of S if I assumed that falsity of presupposition entails non-truth of S (that is, ¬S) and non-truth of ¬S (that is, the falsity of S), and that ¬S, in these two relations, stands for the *same sentence*. If this were my assumption, my analysis would be perfectly Strawsonian: in order to avoid the contradiction between non-truth of S, that is, ¬S, and non-truth of ¬S, it is supposed that falsity of presupposition entails lack of truth-value of S. However, this is not what I am assuming. As argued in (20), I contend that ¬S in the above-mentioned relations is the surface realization of *two different sentences*, which means that falsity of presupposition entails non-truth of S (that is, ¬S$_{(WIDE)}$) and non-truth of ¬S$_{(NARROW)}$. This analysis does not lead to contradiction and is not compatible with an analysis that needs to postulate a lack of truth-value.

(26) Beatrix did not kill Bill.
    a. There was an act of killing Bill such that Beatrix did not perform it.
    b. There was no act of killing Bill such that Beatrix performed it.

If standard entailments may project under negation, how can projection tests be invoked to lay the foundations for a distinction between entailments and presuppositions?[19]

7 The case of factivity

The claim put forward in this paper is that most standard cases of presupposition are classical entailments. Moreover, I claim that all presuppositions that are classical entailments are also pragmatic presuppositions à la Stalnaker, while not all pragmatic presuppositions are also classical entailments. I argue that *factivity* offers a paradigmatic example in this sense. *Factivity* is the term used to denote the relation between a sentence containing a factive verb and the truth of the embedded proposition. Paul and Carol Kiparsky introduced the notion of factivity in their seminal work from 1970 by distinguishing between two types of predicates which take sentences as their complements, namely "factive " (for example, *know/be aware (of), ignore, comprehend, learn, realize, discover, regret, resent, deplore*, etc.) and "non-factive" (for example, *believe, assume, suppose, claim, conclud*e, etc.). They argued that the choice of complement type is especially predictable from the "ᴘʀᴇsᴜᴘᴘosɪᴛɪon by the speaker that the complement of the sentence expresses a true proposition" (Kiparsky & Kiparsky 1970**,** p. 143). Thus, factivity, as originally defined by Kiparsky and Kiparsky, constitutes a case of *speaker presupposition* (in the sense of Stalnaker 1972).

Crucially, if factivity is only considered in terms of speaker presupposition, hence as a pragmatic implication, there is no reason to distinguish between the factivity related to different kinds of factive verbs, namely cognitive verbs (such as *know*) and emotive verbs (such as *regret*). In both cases, we can say that by uttering a factive sentence, the speaker is expressing (or is acting *as if* expressing) her commitment to the truth of the embedded *p*. The account presented in this paper shows that there are reasons for distinguishing between cognitive factives and emotive factives. Indeed, it will be shown that only the implication triggered

---

19 In section 9, we will see that several scholars, in recent works, argue that projection tests are not a diagnostic for "presuppositionality", but rather they ascertain "not-at-issueness".

by *know* is a classical entailment, while the implication triggered by *regret* is a merely pragmatic phenomenon, that is, a speaker presupposition (see, in this sense, Colonna Dahlman 2016, for experimental evidence in favour of this hypothesis, see Djärv *et al.* 2016).

This claim stands in contrast to Karttunen's well-known distinction between *true factives* and *semifactives*. Founding his analysis on the premise that presuppositions should be constant under negation, Lauri Karttunen (1971) points out that *some* factive presuppositions may vanish, rather than being constant, in some environments. Karttunen shows that negative environments are not problematic: it is argued that all the examples in (27) ((22) in Karttunen's text) presuppose the truth of the embedded proposition.

(27) (a)   regret
     (b) John didn't realize that he had not told the truth.
     (c)   discover

On the other side, it is argued that in questions and conditionals factive verbs behave differently. In (28) ((24) in Karttunen's text), it is shown that the factive presupposition goes through in (a) and (b), but not in (c), while in (29) ((25) in Karttunen's text), the factive presupposition survives in (a), but neither in (b) or in (c):[20]

(28) (a)   regret
     (b) Did you realize that you had not told the truth?
     (c)   discover
(29) (a)   regret
     (b) If I realize later that I have not told the truth,
     (c)   discover
      I will confess it to everyone.

Hence the distinction between *true factives*, such as *regret*, that do not lose their factivity in questions and conditionals, and *semifactives*, such as *discover* and *realize* and other similar verbs (*find out, notice, see*), which lose their factivity in questions and conditionals. Karttunen claims that true factives are such that the

---

20 See Stalnaker's (1999, p. 56 ff.) objection: this different behaviour of cognitive factives such as *realize* and *discover* is expected due to the use of the first and second person. Beaver (2010) shows recently that the vanishing of the factive presupposition is not a matter of grammatical person, but rather of information structure.

mere possibility that the antecedent (the matrix clause in which the factive predicate occurs) is true *necessarily* implies the consequent (the complement clause embedded under the factive): if it is possible that X FACTV that *p*, then it must be true that *p*. This relation, according to Karttunen (1971, p. 60), seems to hold for *regret*, as illustrated in (30):

(30) True factives (Karttunen 1971)
      (∀x) (∀s) [M (regret (x, s)) ⇒ s]
      (∀x) (∀s) [M (~regret (x, s)) ⇒ s]

In this line of reasoning, if it is possible that x regrets that s, then s must be true ('M' is a possibility operator). In other words, from the fact that someone may regret that *p*, we must conclude, according to Karttunen, that *p* is the case. By contrast, from the fact that someone may discover or realize that *p* we cannot conclude by necessity that *p* is in fact the case.[21] More extensively, Karttunen's argument goes as follows: given (31), from (31a) one can infer the truth of the complement, whereas there is no such necessary relationship in (31b) and (31c) below.

(31) a. It is possible that I will regret later that I have not told the truth.
      b. It is possible that I will realize later that I have not told the truth.
      c. It is possible that I will discover later that I have not told the truth.

I argue that Karttunen's analysis of factive verbs needs to be reviewed for (at least) two reasons.

---

21 Note that this argument does not apply to *know*. If someone possibly knows that *p*, then *p* must be the case. As a matter of fact, Karttunen never refers to *know* as a semifactive. Indeed, Gazdar (1979a, p. 122, fn 12) mentions the fact that Karttunen (in a personal communication, 1977) "thinks that cognitive factives like *know* do entail their complements, but he is agnostic with respect to emotive factives like *regret*." However, following the classifications proposed by Hooper and Thompson (1973), and by Hooper (1975), *know* is generally considered, in the literature as a semifactive in a Karttunenian sense.

8 Reviewing Karttunen 1971 – Reason 1: *Regret* does not entail *p*

Karttunen's distinction between true factives and semifactives is based on the assumption that if someone regrets that *p*, then *p must* be true. This claim is not tenable. Klein (1975) offered empirical evidence against the assumption that emotive factives necessarily imply the truth of their embedded proposition:[22]

(32) Falsely believing that he had inflicted a fatal wound, Oedipus regretted killing the stranger on the road to Thebes. (Klein 1975, p. B12)

(33) Mary, who was under the illusion that it was Sunday, was glad that she could have a long lie-in. (Klein 1975, p. B12)

In these cases, it is clear that the emotive factives *regret* and *be glad* do not entail the truth of the embedded proposition: both (32) and (33) can be true despite the embedded proposition being false.

In a contrary sense, Gazdar (1979a) argued against Klein's hypothesis that emotive factives like *regret* do not entail the proposition expressed by their complement (Klein 1975, p. B12). His argument is given as follows:

> What seems to be happening in such sentences is that the verb of propositional attitude in the subordinate clause delimits a restricted set of worlds (not including the actual world), and it is only with respect to this set of worlds that the verb of propositional attitude in the matrix sentence gets evaluated.
>
> (Gazdar 1979a, p. 122)

This argument suggests that, in the cases at issue (here examples (32) and (33)), the propositional attitude of regret/being glad as expressed by the matrix verb is under the scope of the propositional attitude of belief expressed by the subordinate clause ('believing that', 'was under the illusion that'). As Gazdar (1979a, p. 123) observes, "The strange thing about these sentences is that material which is syntactically subordinate appears to be semantically superordinate". However, if

---

22 In a similar line of reasoning, see Delacruz (1976, p. 195, fn 8). Other prominent scholars, such as Schlenker (2005) and Égré (2009), have provided further empirical evidence against the assumption that emotive factives entail their embedded proposition. More recently, see Marmor (2014, p. 40, fn 12). Even Karttunen, in a recent paper, admits that "It is clear that we can have false beliefs about the actual world and true regrets based on those false beliefs [..]" (Karttunen 2016, p. 712).

this were the case, then we should assume that the regret/gladness described in these sentences is not a real regret/gladness (a regret/gladness in the actual world), but a belief of regret/gladness. This conclusion is evidently wrong.

More recently, Abrusán (2011, p. 514) follows Gazdar and argues that the issues at hand are cases of *Free Indirect Speech*: they simply report "an attitude of a subject towards facts as perceived by him." That is, the implications related to these sentences do not have to be shared by the speaker, "which accounts for the lack of veridicality" (Abrusán, *ibid.*, see also Abrusán 2016, p. footnote 8, referring to Holton's *Protagonist Projection Hypothesis*[23]). This argument is weak. The sentences in (32) and (33) are not a case of Free Indirect Speech, as standardly defined in the literature (cf. Pascal 1977, Banfield 1982, Prince 2003, Schlenker 2004 and, most recently, Giorgi 2010, Eckardt 2015 and Reboul *et al.* 2016). In fact, the sentences in (32) and (33) would need a serious makeover in order to count as an occurrence of Free Indirect Speech. Here is how they should be transformed in order to be realized as Free Indirect Speech:

(32′) How terrible! He had killed the stranger on the road to Thebes! (Oedipus was not aware that he had not inflicted a fatal wound.)

(33′) Finally it was Sunday – Mary thought – and she could have a long lie-in! (She had not realized yet that it was not Sunday.)

Note that my argument against Abrusán (2011) is not an argument on the correct definition of Free Indirect Speech. Let us assume that Abrusán is right and that the lack of veridicality in (32) and (33) is only due to the fact that the speaker is reporting someone's attitude without committing herself to the truth of the proposition embedded under the attitude verb. If this were correct, it would be possible to build similar structures with other veridical verbs of propositional attitude. This is not the case, and any attempt in this sense would result in semantically unacceptable sentences. Let us take *know* and *remember* (to do something).

---

23 The Protagonist Projection hypothesis was originally proposed by Richard Holton (1997) in order to explain non-factive uses of *know*. According to this hypothesis, *know* is semantically factive, but may be used non-literally (hence non-factively): in this case, it is assumed that the speaker projects herself into the protagonist's state of mind and assumes his/her perspective. The Protagonist Projection hypothesis has been endorsed by Stokke (2013) and by Buckwalter (2014), while it has been criticized by Tsohatzidis (1997, 2012) and by Colonna Dahlman (2017a).

(34) #Falsely believing that Bill was dead, she knew that he had killed him.[24]

(35) #Falsely believing that there was a window to be opened in that room, she
     remembered to open it.

These sentences are clearly anomalous: they cannot be true, and it is not possible
to justify the lack of veridicality by assuming that the speaker is just reporting the
attitude of someone else without being committed to the truth of the embedded
proposition.

Crucially, both the state of affairs in (34) and that in (35) might be described
without contradiction by means of Free Indirect Speech:

(34′) Yes, she was the one who had killed Bill! Now she could tell the news to
     everybody. She had not discovered yet that Bill was alive.

(35′) Right, the window should be opened – she was remembering now – and, of
     course, she would open it. But there was no window to open in that room.

Gazdar's and Abrusán's attempts to criticize the claim put forward by Klein,
namely that emotive factives such as *regret* do not entail their complement, re-
main unsuccessful. *Regret* does not imply the truth of its complement by neces-
sity – it is not veridical.

On the other hand, *know* entails the truth of the proposition expressed by its
complement – it is veridical. Following Wilson (1972), it seems quite straightfor-
ward that (36) is necessarily true:

(36) If John doesn't live in Italy, then no one knows that he lives in Italy.

---

24 The symbol # stands for "semantically inacceptable". Someone could claim that (34) is not
contradictory. However, this is possible only if we interpret the predicate "knew" as "was con-
vinced", that is, as expressing subjective certainty as opposed to epistemic certainty (on the dis-
tinction between subjective certainty [= someone is certain that *p*] and epistemic certainty [= *p*
is certain], see Stanley 2008). In other words, we can assume that *know* is ambiguous between a
factive sense that entails *p* and a non-factive sense that does not entail *p* (see, in this sense, Tso-
hatzidis 2012, Turri 2011, Colonna Dahlman 2016, 2017b). A sentence such as (34) cannot be true
unless we interpret "knew" as meaning something other than ascribing knowledge. On the other
hand, in order for a sentence such as (32) to be true, there is no need to interpret "regretted" as
meaning something other than ascribing regret. This argument is developed in Colonna Dahl-
man (2016).

That is, (37a) entails (37b):

(37) a. Someone knows that John lives in Italy.
      b. John lives in Italy.

On the other hand, a similar line of reasoning does not work for *regret*. A sentence like (38) is not necessarily true:

(38) If John doesn't live in Italy, then no one regrets that he lives in Italy.

In fact, as shown in the above examples, it can be the case that someone who falsely believes that John lives in Italy regrets that John lives in Italy. That is, (39a) does not entail (39b):

(39) a. Someone regrets that John lives in Italy.
      b. John lives in Italy.

In other words, if someone (X) regrets that p, p is not necessarily true, as X's regret may be grounded on a false belief.

Further evidence for a non-uniform account of the implications related to *know* and *regret* is offered in the following conversational scenarios:[25]

(40) Scenario: The (non-)knower
      A: Hi! I know that you live in Italy now!
      B: Well, no, I don't live in Italy.
      A: Oh... Then I never knew it!
      (#Oh... Then I can stop knowing it!)
(41) Scenario: The regretter
      A: I really regret that you now live in Italy!!
      B: But I don't live in Italy!
      A: Oh...Then I can stop regretting it!
      (#Oh... Then I never regretted it!)

If Karttunen's analysis were correct, it would be contradictory to utter something like the sentences in (42):

---

25 I am indebted to Derek Ball, who provided me these examples.

(42) a. I believe that I told the truth, but it is possible that tomorrow I'll regret
   that I didn't do it.
   b. I am giving you this version of the story. I don't know whether it is the
   truth, but I know it's possible that tomorrow I'll regret that I've not been
   telling you the truth.
   c.I lied, but it is possible that one day I'll regret that I actually told the truth.

Contrary to what was claimed by Karttunen concerning the sentences in (31), the
sentences in (42) show that there is no necessary relationship between a *regret*-
sentence and its complement. I contend that a uniform account of factivity with
respect to *know* and to *regret* is uncalled for, as factivity, in the cases of *know* and
regret, has to be intended in two different senses, a semantic one in the case of
know, and a pragmatic one in the case of *regret*. In fact, in the case of *know*, fac-
tivity, on sentence meaning level, is a classical entailment, as the relation be-
tween the sentence containing 'knows' (S) and the embedded proposition (*p*) is a
truth-conditional relation: if S, then necessarily *p*, hence *p* is a truth-condition of
S. On the other hand, in the case of *regret*, the relation between S and *p* is not
truth-conditional, as S can be true even though *p* is false, hence *p* is not a truth-
condition of S. This is because one cannot have knowledge of something false,
while it is perfectly possible to have a regret founded on a false belief.

9 Reviewing Karttunen 1971 – Reason 2: Projective meaning is a diagnostic for backgroundedness

Karttunen's distinction between true factives and semifactives was grounded on
projection tests, that is, tests that were assumed to verify whether alleged presup-
positions project to embedded sentences or not.

In most recent works, it has been argued that projection tests address infor-
mation structure level and are a diagnostic for the discourse status of implica-
tions rather than for presuppositionality. In these works (Potts 2005, Chemla
2009, Tonhauser *et al.* 2013, Abbott 2006, 2016, Simons *et al.* 2017), it is shown
that *not everything which projects is a standard, classical presupposition*. For in-
stance, Chemla (2009) shows that scalar implicatures may behave as presuppo-
sitions – they project in negative environments. Tonhauser *et al.* (2013) argue that
projective content includes several cases of implications, not only cases stand-

ardly described as presuppositions. Abbott (2006, 2016) shows that cases of conventional implicature, such as the implications triggered by connectives (e.g. *therefore, but*), implicative verbs (e.g. *manage, forget to*), iteratives (e.g. *still, again, too*), honorifics (e.g. *tu/vous*), expletives (e.g. *that bastard*), and parentheticals, can project. Simons *et al.* (2010, p. 309) observe that "many expression types yield projection even though they do not have standard properties of presupposition". These scholars put forward the following hypotheses about what projects and why: (i) All and only those implications of (embedded) sentences which are not-at-issue relative to the Question Under Discussion (QUD) in the context have the potential to project; (ii) Operators (modals, negation, etc.) target at-issue content (Simons *et al.* 2010, p. 315). Thus the claim is that material projects if and only if it is actually *not-at-issue*.[26] More recently, the same authors (Simons *et al.* 2017) offer an account of the projection behaviour of the complements of factive verbs within a QUD framework. In particular, they demonstrate that projection of factivity is affected by *focus marking*, which is considered a conventional indicator of the QUD, and by explicit questions. Moreover, and most importantly, they argue that projection may occur with non-factive predicates (such as *believe*), too (Simons *et al.* 2017, p. 203).

All these studies indicate that projection is not a consequence of the 'presuppositional status' of the relevant implication, but is rather a diagnostic for *backgroundedness*, a property that can apply to any kind of implication: if the implication conveys backgrounded information, it projects. In other words, *any implicational content may project depending on information structure factors*. Previously, I have shown that implications standardly considered to be classical entailments may project in negative contexts. As already seen, it is traditionally claimed that a sentence such as 'Beatrix killed Bill' entails 'Bill is dead', while its negative counterpart 'Beatrix did not kill Bill' does not. However, this is not necessarily the case, as we can think of a context that entails 'Bill is dead', where it is true that Beatrix did not kill Bill. For instance, the sentence 'Beatrix did not kill Bill' can be interpreted in the sense that 'There was an act of killing Bill such that Beatrix did not do it, but someone else did it', or in the sense that 'It was not Beatrix who killed Bill: he died of natural causes'. In these cases the entailment triggered by the positive sentence ('Bill is dead') projects under negation. Arguably, this is because the entailment in these cases is not-at-issue content.

---

26 This idea of "not-at-issue content" seems to correspond to Stalnaker's shared knowledge or common ground information (see Stalnaker 1974, 2002) or to Grice's non-controversial propositions, that is, propositions that are not likely to be open to challenge (Grice 1989, p. 274).

Karttunen's projective tests on factives just show that it is easier to embed propositions expressing at issue content – content that is under discussion and might be challenged in the conversation – under cognitive factives than under emotive factives.[27] Why this is the case is another crucial question. My hypothesis is that *regret* tends to embed backgrounded information because of its semantic complexity, following a proposal put forward by Erteschik-Shir for verbs of saying (1973, p. 51–52).[28] The generalization would be that *the more semantically heavy an embedding predicate is, the more difficult it is for this predicate to embed asserted content*. However, I must leave this hypothesis, and the definition of *regret*'s semantic complexity, to future research.[29]

10 Conclusion

This paper proposes a revision of the semantic notion of presupposition. Following Russell (1905), and against the traditional Strawsonian notion of presupposition, I argue that presupposition failure does not result in a lack of truth-value, but rather in falsity, hence that *Modus Tollens* applies to presuppositions and most standard cases of presupposition are classical entailments. Moreover, I claim that all presuppositions that are classical entailments are also pragmatic presuppositions, while not all pragmatic presuppositions are also classical entailments. I contend that factive verbs offer a paradigmatic example of this distinction, and that a uniform account of factivity with respect to *know* and to *regret* is not called for. The term factivity, in the cases of *know* and *regret*, has to be intended in two different senses: a semantic one in the case of *know*, and a pragmatic one in the case of *regret*. If factivity is considered as speaker presupposition, there is no reason to distinguish between the factivity related to *know* and the factivity related to *regret*: in both cases, we can say that by uttering a factive

---

27 This was also argued by Hooper (1975), who classified cognitive factives as *assertives*: the proposition embedded under cognitive factives may be semantically dominant in the sense of Erteschik-Shir (1973), and hence constitute the *main point of assertion* (in this sense, Simons 2007). The weakness of this analysis, however, is that it overlooks the possibility of embedding at-issue content under *regret*.

28 The argument developed by Erteschik-Shir (1973) concerns extraction phenomena with verbs of saying in English. Her hypothesis is that extraction is more acceptable when the verb of saying is semantically less complex and does not imply much about the manner of saying.

29 For a tentative analysis of *regret*'s semantic complexity, see Paper V, section 6, in Colonna Dahlman (2015).

sentence, the speaker is expressing (or is acting *as if* expressing) her commitment to the truth of *p*. The account presented in this paper shows that there are reasons for distinguishing between cognitive factives and emotive factives: only the implication triggered by *know* is a classical entailment, while the implication triggered by *regret* is merely pragmatic. In fact, in the case of *know*, the relation between the sentence containing 'knows' (S) and the embedded proposition (*p*) is a truth-conditional relation: if S, then necessarily *p*, hence *p* is a truth-condition of S. On the other hand, in the case of *regret*, the relation between S and *p* is not truth-conditional, as S can be true even though *p* is false (hence *p* is not a truth-condition of S). This is because one cannot have knowledge of something false, while it is perfectly possible to have a regret founded on a false belief.

It has been argued that Karttunen's traditional distinction between true factives and semifactives is fallacious for two reasons: (i) the distinction is grounded on a wrong assumption, which is that *regret* entails its complement; (ii) the distinction is based on projection tests, which are not a diagnostic for presuppositionality, but rather only show that it is easier to embed at-issue content under cognitive factives than under emotive factives. These tests tell nothing about the semantic relation between the factive sentence and its embedded proposition.

Acknowledgments

References

Abbott, Barbara (2000): "Presuppositions as Nonassertions". In: *Journal of Pragmatics* 32, pp. 1419–1437.

Abbott, Barbara (2008): "Presuppositions and Common Ground". In: *Linguistics and Philosophy* 31, pp. 523–538.

Abbott, Barbara (2016): "An Information Packaging Approach to Presuppositions and Conventional Implicatures". In: *Topoi* 35, pp. 9–21.

Abrusán, Márta (2011): "Predicting the Presuppositions of Soft Triggers". In: *Linguistics and Philosophy* 34, pp. 491–535

Abrusán, Márta (2016): "Presupposition Cancellation: Explaining the 'Soft-Hard' Trigger Distinction". In: *Natural Language Semantics* 24, pp. 165–202.

Austin, John L. (1962): How to Do Things with Words. The William James Lectures delivered at Harvard University in 1955. Oxford/New York: Oxford University Press.

Banfield, Ann (1982): Unspeakable Sentences. Narration and Representation in the Language of Fiction. Boston/London: Routledge & Kegan Paul.

Beaver, David (2010): "Have you Noticed that Your Belly Button Lint Colour is Related to the Colour of Your Clothing?" In: Rainer Bäuerle/Uwe Reyle/Thomas Ede Zimmermann (eds.): *Presuppositions and Discourse: Essays offered to Hans Kamp*. Bingley: Emerald Group Publishing Limited, pp. 65–99.

Beaver, David I., and Bart Geurts (2013): "Presupposition," *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.). Available at: http://plato.stanford.edu/archives/fall2013/entries/presupposition/

Bezuidenhout, Anne (2016): "Presupposition Failure and the Assertive Enterprise". In: *Topoi* 35(1), pp. 23–35.

Buckwalter, Wesley (2014): "Factive Verbs and Protagonist Projection". In: *Episteme* 11(4), pp. 391–409.

Casalegno, Paolo (1998): *Filosofia del Linguaggio*. Roma: Carocci editore.

Chemla, Emmanuel (2009): "An Anti-Introduction to Presupposition". In: Paul Égré/Giorgio Magri (Eds.): *Presuppositions and Implicatures: Proceedings of the MIT-Paris Workshop*. MIT Working Papers in Linguistics 60, pp. 35–46.

Chierchia, Gennaro, and Sally McDonnell-Ginet (1990): *Meaning and Grammar*. Cambridge, MA: The MIT Press.

Colonna Dahlman, Roberta (2015): Studies on Factivity, Complementation, and Propositional Attitudes. Ph.D. Thesis, Lund University.

Colonna Dahlman, Roberta (2016): "Did People in the Middle Ages *Know* that the Earth was Flat?" In: *Acta Analytica* 31(2), pp. 139–152.

Colonna Dahlman, Roberta (2017a): "The Protagonist Projection Hypothesis. Do We Need It?" In: *International Journal of Pragmatics* 9, pp. 134–153.

Colonna Dahlman, Roberta (2017b): "The Ambiguity of *Sapere*". In: *Italian Journal of Linguistics* 29(2): 3–28.

Delacruz, Enrique B. (1976): "Factives and Proposition Level Constructions in Montague Grammar". In: Barbara H. Partee (Ed.): *Montague Grammar*. New York: Academic Press, 177–199.

Dinsmore, John (1981): "Towards a Unified Theory of Presupposition". In: *Journal of Pragmatics* 5: 335–363.

Djärv, Kajsa, Zehr, Jérémy, and Florian Schwarz (2016): "Cognitive vs. Emotive Factives: An Experimental Differentiation". In *Proceedings of Sinn und Bedeutung 21*, Edinburgh, September 4-6, 2016. Available at: https://sites.google.com/site/sinnundbedeutung21/proceedings-preprints

Eckardt, Regine (2015): The Semantics of Free Indirect Discourse. How Texts Allow Us to Mind-Read and Eavesdrop. Leiden/Boston: BRILL.

Égré, Paul (2009): "Question-Embedding and Factivity". In: Paul Égré/Giorgio Magri (Eds.): *Presuppositions and Implicatures: Proceedings of the MIT-Paris Workshop*. MIT Working Papers in Linguistics 60, pp. 79–99.

Erteschik-Shir, Nomi (1973): *On the Nature of Island Constraints*. Ph.D. Thesis, M.I.T.

Erteschik-Shir, Nomi (1997): *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.

Frege, Gottlob (1892): "Über Sinn und Bedeutung". In: *Zeitschrift für Philosophie und Philosophische Kritik*, NF 100, pp. 25–50.

Frege, Gottlob (1948): "Sense and Reference". In: *The Philosophical Review* 57(3), pp. 209–230.

Gauker, Christopher (2008): "Against Accommodation: Heim, van der Sandt, and the Presupposition Projection Problem". In: *Philosophical Perspectives* 22, pp. 171–205.

Gauker, Christopher (2016): "Presuppositions as Anaphoric Duality Enablers". In: *Topoi* 35(1), pp. 133–144.

Gazdar, Gerald (1979a): *Pragmatics: Implicature, Presupposition, and Logical Form*. New York: Academic Press.

Gazdar, Gerald (1979b): "A Solution to the Projection Problem". In: Choon–Kyu Oh/David A. Dinneen (Eds.): *Syntax and Semantics, Vol. 11: Presupposition*. New York: Academic Press, pp. 57–89.

Geach, Peter T. (1950): "Russell's Theory of Descriptions". In: *Analysis* 10(4), pp. 84–88.

Giorgi, Alessandra (2010): *About the Speaker: Towards a Syntax of Indexicality*. Oxford: Oxford University Press.

Grice, Paul (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Hintikka, Jaakko (1962): *Knowledge and Belief. An Introduction to the Logic of the Two Notions*. Ithaca/New York: Cornell University Press.

Holton, Richard (1997): "Some Telling Examples: A Reply to Tsohatzidis". In: *Journal of Pragmatics* 28, pp. 625–628.

Hooper, Joan B./ Thompson, Sandra A. (1973): "On the Applicability of Root Transformations". In: *Linguistic Inquiry* 4(4), pp. 465–497.

Hooper, Joan B. (1975): "On Assertive Predicates". In: John P. Kimball (Ed.): *Syntax and Semantics, Vol. 4*. New York: Academic Press, pp. 91–124.

Huang, Yan (2007): *Pragmatics*. Oxford: Oxford University Press.

Karttunen, Lauri (1970): "On the Semantics of Complement Sentences". In: *Papers from the Sixth Regional Meeting*. Chicago Linguistic Society, April 16–18, 1970, pp. 328–339.

Karttunen, Lauri (1971): "Some Observations on Factivity". In: *Papers in Linguistics* 4(1), pp. 55–69.

Karttunen, Lauri (1973): "Presuppositions of Compound Sentences". In: *Linguistic Inquiry* 4(2), pp. 169–193.

Karttunen, Lauri (1974): "Presupposition and Linguistic Context". In: *Theoretical Linguistics* Vol.1, 1(2), pp. 181–194.

Karttunen, Lauri (2016): "Presupposition: What Went Wrong?" In: Mary Moroney/Carol-Rose Little/Jacob Collard/Dag Burgdorf (Eds.): *Proceedings of Semantics and Linguistic Theory 26*. Linguistic Society of America and Cornell Linguistics Circle, pp. 705–731.

Keenan, Edward L. (1971): "Two Kinds of Presupposition in Natural Language". In: Charles J. Fillmore/D. Terence Langendoen (Eds.): *Studies in Linguistic Semantics*. New York: Holt, Rinehart and Winston, Inc., pp. 45–52.

Kiparsky, Paul/Kiparsky, Carol (1970): "Fact". In: Manfred Bierwisch/Karl Erich Heidolph (Eds.): *Progress in Linguistics: A Collection of Papers*. The Hague/Paris: Mouton, pp. 143–173.

Klein, Ewan (1975): "Two Sorts of Factive Predicate". In: *Pragmatics Microfiche* Vol. 1, Fiche 1, pp. B5–C14.

Langendoen, D. Terence/Harris B. Savin. (1971): "The Projection Problem for Presuppositions". In: Charles J. Fillmore/D. Terence Langendoen (Eds.): *Studies in Linguistic Semantics*. New York: Holt, Rinehart and Winston, Inc., pp. 54–60.

Levinson, Stephen C. (1983): *Pragmatics*. Cambridge: Cambridge University Press.

Macagno, Fabrizio/Alessandro Capone. (2016): "Presuppositions as Cancellable Inferences". In: Keith Allan/Alessandro Capone/Istvan Kecskes (Eds.): *Pragmemes and Theories of Language Use*. Perspectives in Pragmatics, Philosophy & Psychology 9. Springer International Publishing Switzerland, pp. 45–68.

Marmor, Andrei (2014): *The Language of Law*. Oxford: Oxford University Press.

Neale, Stephen (1990): *Descriptions*. Cambridge, MA: The MIT Press.

Pascal, Roy (1977): *The Dual Voice*. Manchester: Manchester University Press.

Potts, Christopher (2005): *The Logic of Conventional Implicature*. Oxford: Oxford University Press.

Prince, Gerald (2003): *A Dictionary of Narratology*. Lincoln/London: University of Nebraska Press.

Quine, Willard van Orman (1960): *Word and Object*. Cambridge, MA: The MIT Press.

Reboul, Anne/Denis Delfitto/Gaetano Fiorin (2016): "The Semantic Properties of Free Indirect Discourse". In: *Annual Review of Linguistics* 2, pp. 255–271.

Reinhart, Tanya (1981): "Pragmatics and Linguistics: An Analysis of Sentence Topics". In: *Philosophica* 27(1), pp. 53–93.

Russell, Bertrand (1905): "On Denoting". In: *Mind* 14(56), pp. 479–493.

Sbisà, Marina (1999): "Presupposition, Implicature and Context in Text Understanding". In: Paolo Bouquet/Massimo Benerecetti/Luciano Serafini/Patrick Brézillon/Francesca Castellani (Eds.): *Modeling and Using Context*. CONTEXT 1999. Lecture Notes in Computer Science, vol. 1688. Berlin/Heidelberg: Springer, pp. 324–338.

Schlenker, Philippe (2004): "Context of Thought and Context of Utterance: A Note on Free Indirect Discourse and the Historical Present". In: *Mind and Language* 9(3), pp. 279–304.

Schlenker, Philippe (2005): "The Lazy Frenchman's Approach to the Subjunctive: Speculations on Reference to Worlds and Semantic Defaults in the Analysis of Mood". In: Twan Geerts/Ivo van Ginneken/Haike Jacobs (Eds.): *Romance Languages and Linguistic Theory. Selected Papers from 'Going Romance' 2003, Nijmegen, 20-22 November*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 269–309.

Sellars, Wilfrid (1954): "Presupposing". In: *The Philosophical Review* 63(2), pp. 197–215.

Seuren, Pieter A.M. (2006): "Presupposition". In: *Encyclopedia of Language and Linguistics*, pp. 80–87.

Simons, Mandy (2001): "On the Conversational Basis of Some Presuppositions". In: Rachel Hastings/Brendan Jackson/Zsofia Zvolenszky (Eds.): *Proceedings of Semantics and Linguistic Theory 11*. Ithaca, NY: CLC Publications, pp. 431–448.

Simons, Mandy (2006): "Presupposition Without Common Ground". *PhilPapers*. Available at: http://www.hss.cmu.edu/philosophy/simons/Presupposition%20without%20Common%20Ground.pdf

Simons, Mandy (2007): "Observations on Embedding Verbs, Evidentiality, and Presupposition". In: *Lingua* 117, pp. 1034–1056.

Simons, Mandy/Judith Tonhauser/David Beaver/Craige Roberts (2010): "What Projects and Why". In: Nan Li/David Lutz (Eds.): *Proceedings of Semantics and Linguistic Theory 20*. Ithaca, NY: CLC Publications, pp. 309–327.

Simons, Mandy (2013): "Presupposing". In: Marina Sbisà/Ken Turner (Eds.): *Pragmatics of Speech Actions*. Berlin: Mouton De Gruyter, pp. 143–172.

Simons, Mandy/Beaver, David/Roberts, Craige/Judith Tonhauser (2017): "The Best Question: Explaining the Projective Behavior of Factives". *Discourse Processes* 54(3), pp. 187–206.

Soames, Scott (1989): "Presupposition". In: Dov M.Gabbay/Franz Guenthner (Eds.): *Handbook of Philosophical Logic*, Vol. 4. Dordrecht: Reidel, pp. 553–616.

Stalnaker, Robert (1972): "Pragmatics". In: Donald Davidson/Gilbert Harman (Eds.): *Semantics of Natural Language*. Dordrecht/Boston: D. Reidel Publishing Company, pp. 380–397.

Stalnaker, Robert (1973): "Presuppositions". In: *Journal of Philosophical Logic* 2, pp. 447–457.

Stalnaker, Robert (1974): "Pragmatic Presuppositions". In: Milton K. Munitz/Peter K. Unger (Eds.): *Semantics and Philosophy*. New York: New York University Press, pp. 197–213.

Stalnaker, Robert (1984): *Inquiry*. Cambridge, MA: The MIT Press.

Stalnaker, Robert (1999): *Context and Content. Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.

Stalnaker, Robert (2002): "Common Ground". In: *Linguistics and Philosophy* 25, pp. 701–721.

Stanley, Jason (2008): "Knowledge and Certainty". In: *Philosophical Issues* 18, pp. 35–57.

Stanosz, Barbara (1991): *10 Wykładów z Filozofii Języka*. Warszawa: Polskie Towarzystwo Semiotyczne.

Stokke, Andreas (2013): "Protagonist Projection". In: *Mind & Language* 28(2), pp. 204–232.

Strawson, Peter F. (1950): "On Referring". In: *Mind*, New Series 59(235), pp. 320–344.

Strawson, Peter F. (1952): *Introduction to Logical Theory*. London: Methuen & Co.

Strawson, Peter F. (1964): "Identifying Reference and Truth-Values". In: *Theoria* 30, pp. 96–118.

Tonhauser, Judith/Beaver, David/Roberts, Craige/ Simons, Mandy (2013): "Toward a Taxonomy of Projective Content". In: *Language* 89(1), pp. 66–109.

Tsohatzidis, Savas L. (1997): "More Telling Examples: A Response to Holton". In: *Journal of Pragmatics* 28, pp. 629–636.

Tsohatzidis, Savas L. (2012): "How to Forget that 'Know' is Factive". In: *Acta Analytica* 27, pp. 449–459.

Turner, Ken (1992): "Defending Semantic Presupposition". In: *Journal of Pragmatics* 18, pp. 345–371.

Turri, John (2011): "Mythology of the Factive". In: *Logos & Episteme* 2(1), pp. 143–152.

van Fraassen, Bas C. (1968): "Presupposition, Implication, and Self-Reference". In: *The Journal of Philosophy* 65(5), pp. 136–152.

von Fintel, Kai (2004): "Would You Believe It? The King of France is Back! (Presuppositions and Truth-Value Intuitions)". In: Marga Reimer/Anne Bezuidenhout (Eds.): *Descriptions and Beyond*. Oxford: Oxford University Press, pp. 315–341.
Wilson, Deirdre (1972): "Presuppositions on Factives". In: *Linguistic Inquiry* 3(3), pp. 405–410.

Andrei Moldovan

Can Entailments Be Implicatures?

Abstract: Is it possible for the same proposition to be both entailed and conversationally implicated by what is said? A review of the literature that discusses implicatures reveals that the question in the title has received different answers. While those who answer the question affirmatively rely on examples of entailments that look very much like typical cases of conversational implicatures, those who answer it negatively adduce considerations pertaining to the conceptual difference that needs to be drawn between the two categories of implications. In this chapter, I consider the various criteria for identifying conversational implicatures (CIs for short) and conclude that two features of CIs could offer support for ruling out entailments from the class of contents that could be conversationally implicated: *cancellability* and *non-conventionality*. While the cancellability criterion does not lead to a compelling answer to our question, non-conventionality does. Moreover, non-conventionality is a defining feature of CIs, according to Grice's official definition of CIs (1989, p. 26). So, if we follow Grice strictly on this point, the answer to our question should be negative. However, in section 5 I consider the possibility of revising Grice's definition of CIs. I discuss two reasons given in the literature for taking non-conventionality to be a defining feature of CIs, and I argue that they are not compelling. In consequence, I propose a definition of CI in terms of calculability alone. This definition, I argue, does not rule out entailments. Thus, according to my proposal, an implication has the status of *entailment* if it bears a certain logical relation to the semantic content of the utterance; and it has the status of a *conversational implicature* if it is conversationally calculable. So, the answer to our initial question is, eventually, a tentative 'yes'.

Keywords: entailment, conversational implicature, conventional implicature, cancellability, conventionality, definition, Grice

1 Introduction

The question I address in this paper is formulated in the title: whether an entailment q of the proposition p literally expressed by an utterance of a natural language sentence can, at the same time, figure as the content of a conversational

implicature (CI, henceforth) that the utterance carries.[1] In other words, given that both entailments and conversational implicatures are implications of *what is said*, is it possible for the same proposition to be both entailed and conversationally implicated by what is said? Of course, not all entailments are implicatures, and not all implicatures convey contents that are entailed by what is said. The question is whether this is possible, that is, whether a proposition *could* play this double role.

While there is no sustained discussion in the literature of the question we are concerned with here (or at least I have found none), the first place one might think of in order to search for an answer is Paul H. Grice's seminal papers on implicatures. Although I can find no direct discussion of our question in his articles, they do offer reasons for a particular answer to this question. I consider some of them below (in section 5). But probably the most relevant passage is the following:

> Since the truth of a conversational implicatum is not required by the truth of what is said (what is said may be true – what is implicated may be false), the implicature is not carried by what is said, but only by the saying of what is said, or by "putting it that way".
>
> (Grice 1989, p. 39)

In this passage Grice emphasizes the idea that the relation between conversational implicatures and what is said is not a truth-conditional relation. He arrives at this conclusion by pointing out that the implicatum could be false even if what is said (i.e. the literal content of the utterance) is true, which means that the latter is not entailed by the former. But does Grice mean here that the truth of a CI is *never* entailed by what is said, or that it is *not always* entailed by what is said? If he means the former, then Grice's answer to our initial question is negative. However, Grice might mean in the quoted passage simply that there might be cases in which the implicatum is not entailed by what is said, allowing for the possibility of cases in which it is. If that is what he means, then it is not a condition on the implicatum that it is *not* entailed by the semantic content. This interpretation is favoured by the fact that this claim is sufficient to support the conclusion Grice draws in the quoted passage, that the implicature is not carried by what is said, but by the saying of it. If this is the right interpretation of the passage, then Grice is *not* suggesting here that entailments cannot be the content of implicatures.

Apart from Grice's own view on the matter, it is interesting to consider whether there are good reasons to be found in the literature on implicatures in support of a particular answer to our question. Indeed, several authors have addressed the question we are concerned with here. Some of them find an affirmative answer to the question unproblematic, and simply notice that there are cases in which the content of an implicature is entailed by the semantic content of the utterance. Kent Bach (2006, p. 24) endorses this position: "It is commonly assumed that what a speaker implicates in uttering a sentence can't be entailed by the sentence itself. To be sure, most implicatures (by speakers) are not entailments (by sentences uttered by speakers), but there are exceptions". See also Wayne Davis (1998, p. 6) and Blome-Tillmann (2013, p. 4), Carston (2002, p. 139), Higashimori and Wilson (1996, p. 12) for very similar, and equally quick remarks. Those who adopt this position usually offer examples of such CIs, which they take to prove their case. Here are some such examples (where the entailment relation is symbolized with '⊨' and the relation between the utterance of the sentence and the implicatum with '+>'):

(1) A: Nobody has ever long-jumped over 28 feet.
    B: Whad'ya mean? Bob Beamon long-jumped over 29 feet way back in 1968.
        +> Somebody long-jumped over 28 feet.
        ⊨ Somebody long-jumped over 28 feet. (Bach 2006, p. 5)
(2) A: Does John drink slivovitz?
    B: He doesn't drink any alcohol.
        +> John does not drink slivovitz.
        ⊨ John does not drink slivovitz. (Carston 2002, p. 139)
(3) A: Are you or your spouse 65 or older or blind?
    B: I am 67.
        +> I or my spouse is 65 or older or blind.
        ⊨ I or my spouse is 65 or older or blind. (modified from Davis 1998, p. 6)

Indeed, we have here a *prima facie* reason for a positive answer to the question, which is the following: the contents introduced by the sign '+>' are typical cases CIs, or at least look very much like CIs; at the same time, these contents are entailed by the proposition literally expressed by the utterances of the respective

sentences; therefore, an entailed proposition can at the same time be the content of an implicature.[2]

Are these examples sufficient to prove the case for an affirmative answer to our question? While it seems clear that the content on the last line is indeed entailed by what is said, the other premise of the argument might raise doubts: although the above examples look very much like typical cases of CIs, one might simply reject the claim that they *are* implicatures. Those who give a negative answer to the question provide a reason to do so. As Stephen Neale (1992, p. 18) puts it, "[i]ntuitively, it seems desirable that no proposition be both an entailment and a conversational implicature of the same utterance." The suggestion is that CIs and entailments are different *kinds* of implications of an utterance of a sentence, and any good classification is such that its categories do not intersect. More precisely, proponents of a negative answer to the question under discussion suggest that the concept of implicature and that of entailment need to have a null intersection for reasons having to do with the theoretical role these concepts play. As Michael Haugh (2013, p. 4) notes, "many scholars treat implicatures as essentially synonymous with non-logical/pragmatic inference". He mentions in this sense Carston (2002), Levinson (1983, 2000) and Sperber and Wilson (1995). In the same vein, Laurence Horn (2012, p. 74) writes, "implicature was (re)introduced into the philosophical literature… as a species of implication distinct from logical implication or entailment".

Indeed, these conceptual considerations are especially important in the case of technical terms, and Grice introduces "implicature" as "a term of art" (1989, p. 24), a technical term. While we are familiar with, and sensitive to, the phenomenon of implicatures, as detecting implicatures is part of the normal process of utterance interpretation, the term is not meant to express a concept that we already have, but to introduce a new one. What this indicates is that a better strategy than that of answering the question on the basis of untutored intuitions alone (as the proponents of the affirmative answer do) is to begin by considering the conceptual features we need to ascribe to the notion of CI in order to do the theoretical work it is meant to do, and see whether they rule out entailments.

---

2 An anonymous reviewer disagrees with the claim that these are *prima facie* cases if CI. S/he suggests that what these examples show is that entailed content may be "the main point of assertion" (i.e. the answer to the question under discussion at that stage of the conversation), but not that it is conversationally implicated. Indeed, that an entailed content is the main point of assertion does not mean that it is a CI. Nevertheless, there is an undeniable similarity between examples (1), (2) and (3) and typical cases of CI, on which, I take it, the authors draw when they take the latter to *be* CIs. This intuition is also reinforced by considerations concerning calculability, which I discuss below.

2 Conceptual considerations

The following definition of 'entailment' I take to be satisfactory for our present purposes. Thus, assuming a broadly conceived Kaplanian framework for natural language semantics (see Kaplan 1989), one could define entailment as follows: for any two propositions p and q, p *entails* q iff for any (w, t) such that p(w, t) = 1 it is also the case that q(w, t) = 1. Here w stands for a possible world, t for an instance of time, and 1 and 0 for true and false, respectively.[3]

Notice that, on this definition of entailment, not only "at-issue" entailments such as the ones in examples (1) to (3), but also certain lexically-triggered presuppositions are entailments. For instance, consider the entailments introduced by factive verbs or cleft sentences:[4]

(4) It was John who gave me the address.
     ⊨ Somebody gave me the address.
(5) I realized I had made a mistake.
     ⊨ I had made a mistake.

This inclusive definition of 'entailment' is acceptable for present purposes, as these presuppositions can also feature as possible candidates for the content of CIs, as the following example shows:

(6) A: Did you invite your colleague to the event?
     B: I've just realized I did not invite him.
        +> I did not invite him.
        ⊨ I did not invite him.

Let us now turn to CI. I start by considering the various criteria that have been offered in the literature with the purpose of identifying CIs, consider whether they are necessary properties of CIs, and whether they rule out entailments. Grice

---

3 The appeal to the Kaplanian framework is not essential for the present discussion. Alternatively, one could assume a framework in which propositions receive a truth value relative to a different n-tuple of parameters of evaluations, which features other (or more) parameters than w and t. Also, one could consider defining the entailment relation for pairs of sentences evaluated relative to a context of utterance instead of propositions. I prefer to consider propositions in the above definition given that the implicatum is not a sentence, but a proposition.
4 Potts (2005, p. 23) treats both lexically-triggered presuppositions and at-issue entailments under the category "entailments", in which he also includes conventional implicatures.

(1989, p. 39) lists several such criteria to distinguish between conventional and conversational implicatures: (a) being cancellable; (b) non-detachable; (c) non-conventionally triggered; (d) carried by the utterance of the sentence, and not by what is said; (e) indeterminate. To this we can add (f) calculability, which he mentions later on (1989, p. 44). We should also add (g) being conversational, given that Grice allows for the possibility of implicatures that are non-conventional and non-conversational, and which are generated by "aesthetic, social, or moral" maxims (1989, p. 28). A final candidate is (h) being intended to be conveyed by the speaker. With this we obtain a list of criteria usually used in the literature to determine whether a felt implication is a CI. Let us consider them one by one and see what they tell us about the concept of implicature.

Some of the criteria mentioned above clearly do not rule out entailments. Thus, a CI is (b) non-detachable if it is not possible to find an alternative way of expressing the same (or very similar) literal content without conveying the CI. But, of course, the propositions entailed by what is said are non-detachable, so this criterion can be disregarded.[5] Moreover, non-detachability is not a necessary condition for the presence of a CI, because there are implicatures that depend on the Maxim of Manner (see Grice 1989, p. 37), as well as cases in which there is no alternative way to express the same literal content.

The same can be said of point (h), which refers to the intentional character of CIs. According to Grice, CIs are part of speaker meaning, which is to be analysed in terms of the speaker's communicative intentions. So, it seems (h) is a necessary condition of any CI.[6] However, the entailments that are candidates for CIs are also intended to be conveyed as part of speaker meaning.[7] This criterion does not rule out entailments as candidates for the content of CIs.

Point (e) on our list is indeterminacy. If CIs are necessarily indeterminate, then entailments are ruled out. However, although it is a usual characteristic of

---

5 In fact, Sadock (1978), who subscribes to a negative answer to our question, argues that the test is not sufficient to characterize CIs because it does not distinguish them from entailments: "It is not possible to paraphrase *Bill and Harry left* without conveying *Harry left*. Yet this is clearly not an example of conversational implicature, but of logical inference." (1978, p. 288). Of course, this argument against non-detachability rests on the premise that entailments cannot be the content of CIs, a premise which is under discussion here.

6 Neale (1992, p. 528) writes that (h) follows, if not from the steps of the derivation, "at least from the fact that (a) what [the utterer] U implicates is part of what U means, and (b) what U means is determined by U's communicative intentions."

7 Not all entailments are part of speaker meaning: if I literally say p, I usually do not mean to say $p \vee q \vee r \vee s$.... (I thank Julia Zakkou for this observation.)

CIs, it is not a necessary trait, as Sadock (1978, p. 285) observed. And, as Moeschler (2012, p. 417) notes, it is not difficult to find exceptions, especially for cases of generalized CI. For instance, consider Bach's (1994, p. 139) example of an utterance of the following sentence:

(7)   Jack and Jill are married.

The standard interpretation of (7) is that Jack and Jill are married *to each other*. It looks like there is not more than one possible completion of the propositional content so as to obtain the implicatum.[8]

There are also criteria in the above list that are reducible to others. For instance, (f) calculability does not simply require that there be a possible correct pattern of reasoning that might lead us to the conclusion that the speaker intends to convey q beyond the literal content p. Instead, 'calculability' means something more precise: that there is a particular pattern of reasoning that relies essentially on the assumption that the speaker is observing the Cooperative Principle and the *conversational* maxims (not other maxims). In this sense, if an implication is calculable it is also conversational, and so (f) being calculable and (g) being conversational are essentially the same property. I discuss the criterion of calculability in more detail below.

A similar point can be made about (d), which is the criterion according to which, as Grice (1989, p. 39) puts it, a CI is not carried by what is said, but only by the saying of what is said, or by "putting it that way". That is, being the content of a CI does not consist in bearing a particular relation to a proposition literally expressed itself, but to the utterance of the sentence that has that literal content and which was uttered with specific communicative intention, and in a context in which certain assumptions about the conversation are in place. Although the entailment relation holds between the entailed content and what is said (not the utterance of the sentence), this fact does not by itself conflict with the requirement that the CI is carried by the utterance of the sentence. So, (d) does not exclude entailments from figuring as contents of CIs.

At this point, we are left with: (a) cancellability, (b) non-conventionality, and (f) calculability. I discuss them separately in the next two sections.

---

8 Bach prefers to call this a "conversational impli*ci*ture", the difference with CIs being that in the former case the content conveyed is "closely related" (1994, p. 126) to what is said, being an expansion of the proposition literally expressed, or, in case the literal content is sub-propositional, a completion of it so as to achieve full propositional content. However, his terminological choice is not particularly relevant here.

3 Calculability

According to Grice, calculability is a *necessary* condition for CI: "the final test for the presence of a conversational implicature [has] to be, as far as I [can] see, a derivation of it" (1981, p. 187).[9] Sadock (1978) seconds him on this opinion: "Only one feature, calculability, is clearly a necessary property of conversational implicature."

If we agree to take calculability to be a necessary condition for CIs, then, in order to answer the question of this essay, we need to ask whether an entailment can ever be derivable. Now, the fact that q is entailed by what is said, and, in some sense, part of the speaker meaning (at least in the case of the examples discussed) does not guarantee that there is a derivation of q that makes essential use of the conversational maxims. However, it does not exclude this possibility either. The derivation that we are looking for is an argument which leads to the conclusion that *the speaker intends to convey that q* by way of a pattern of reasoning that appeals to considerations of rationality, information that is shared in the context, and, most importantly, the conversational maxims. To take an example, consider the (alleged) CI carried by B's answer to A's question in example (3). Adapting the derivation scheme that Grice (1989, p. 31) proposes to the present case, the calculation has the following steps (where p is: *B is 67*, and q is: *B or B's spouse is 65 or older or blind*):

  i.   B has said that p;
 ii.   B has apparently violated the maxim of Relevance, yet I have no reason to suppose that she is opting out from the operation of the Cooperative Principle;
iii.   B could not be doing this unless she thought either that q or that ¬q;
 iv.   B knows (and knows that A know that B knows) that A can see that B thinks the supposition that B thinks either that q or that ¬q is required;
  v.   B has said that p, which is incompatible with ¬q; therefore, B does not intend A to think that ¬q;
 vi.   B has done nothing to stop A from thinking that q, and B intends A to think, or is at least willing to allow A to think, that q;
vii.   and so, B has implicated that q.

---

9 See also Grice (1989, pp. 31, 43) for similar remarks. Neale (1990, p. 78) calls this the Justification Requirement.

So, calculability does not rule out entailments from candidates to CIs. Indeed, Neale (1992), who proposes a negative answer to the question whether entailments can be CIs, noted "[e]ntailments do not seem to have been excluded [by the calculation requirement]" (1992, p. 18).

4 Cancellability

Cancellability refers to the success of a potential operation of cancellation, achieved by performing a specific speech act with a particular intention (especially in the case of what Grice calls "explicit cancelation"). Is cancellability a necessary condition of CIs? The received view in pragmatics follows Grice (1989, p. 44) in considering that "all conversational implicatures are cancellable".[10] However, arguments to the contrary have also been advanced. The debate over the reliability of the cancellability test for CIs has recently expanded considerably, with several objections to the test being advanced and discussed in the literature. I briefly mention three of the most discussed objections to the cancellability test for CIs.

One such objection is due to Weiner (2006), who argues that there are cancellation-resistant implicatures, for which explicit cancellation phrases (such as 'But I do not mean to imply that q', where q is the implicatum of the previous sentence uttered) end up reinforcing the implicature, instead of cancelling it. There are various ways in which an attempt to cancel an implicature might go wrong or be infelicitous (see for instance Colonna-Dahlman 2013, and Åkerman 2015). Thus, it is an empirical issue whether CIs are always cancellable or not, and not a conceptual issue.[11] Various responses to this line of argument have been formulated in Blome-Tillmann (2008), Borge (2009), Colonna Dahlman (2013), Åkerman (2015), among others, including reformulations of the cancellability test for CIs so as to avoid the objection.

A second objection to the cancellability test is based on the idea that if CIs are necessarily part of speaker meaning, that is, are intended by the speaker to be conveyed to an audience, then the very idea of cancellation requires postulat-

---

10 Sadock (1978, p. 295) also writes that "Cancellability is probably a necessary feature of conversational implicature, but it gets progressively harder to cancel an implicature the more generalized it is..."
11 Haugh (2013, p. 136) also notes that "[w]hether an implicature can be plausibly or legitimately denied, retracted or clarified by the speaker is... an empirical question".

ing contradictory intentions. This means that, by definition, CIs cannot be cancellable. This argument is proposed in Burton-Roberts (2010) and Capone (2009), but see Haugh (2013, pp. 139–142) for a reply. If either of these two arguments against the cancellability test is sound, then cancellability obviously cannot be a necessary feature of CIs.[12]

A third challenge to the cancellability test has been proposed by various authors, including Wilson and Sperber (1986, p. 61), Carston (2002, p. 140), Blome-Tillmann (2013, p. 4), and others. The objection rests on the assumption that in certain special cases CIs can be entailed by what is said. Given that entailments are not cancellable, the conclusion follows that not all CIs are cancellable. The problem with this argument is that it begs the question that this essay raises: it *assumes* that entailments can be CIs in certain special circumstances. We do not want to take this for granted, but instead look for a reasoned answer to our question. So, this objection to the cancellability test loses its force, at least in this context, and until an answer to the initial question is reached.

Now, if cancellability is taken to be a necessary condition for CIs, this seems to raise a serious problem for those who defend that entailments can be CIs, given that the latter are apparently non-cancellable without contradiction. This is the reason Neale (1992, p. 18) offers for the claim that entailments cannot be implicatures: "Unlike an entailment, a conversational implicature is supposed to be cancellable either explicitly or contextually, without contradiction." Indeed, the attempt to cancel an entailment explicitly appears to result in a contradiction. Consider again example (1), here completed to obtain (8):

(8) A: Nobody has ever long-jumped over 28 feet.
   B: ?? Whad'ya mean? Bob Beamon long-jumped over 29 feet way back in 1968. But I do not mean to imply that someone long-jumped over 29 feet way back in 1968.

It seems there could be no context in which this exchange could take place and B's utterance be interpretable as non-contradictory.

---

12 I do not discuss here these arguments in detail, but only note that, although the debate is still alive, most of the participants so far seem to agree that the objections are not compelling. Given that the default position in the literature on pragmatics takes cancellability to be a necessary property of CIs, I take it that the burden of proof is on those who reject the test, a burden which they do not seem to have met.

However, it has been argued that there are other, more successful, ways of cancelling the entailment. According to Haugh (2013, p. 138), B could have answered in the following way, thus cancelling the implication that *somebody long-jumped over 28 feet*.

(9) B: Whad'ya mean? Bob Beamon long-jumped over 29 feet way back in 1968...
     although, if you mean since Beamon's jump then you're right.

I do not think Haugh's example provides a compelling case of cancellation of an entailment. What seems to be going on here is the following: the additional conditional sentence introduced by "although" provides a new interpretation to A's initial utterance, on which "ever" refers to the interval of time after Beamon's jump in 1968. On this alternative interpretation, the existential proposition that *someone ever long-jumped over 28 feet* does not contradict A's assertion anymore. In fact, B's first assertion loses its conversational relevance, which B acknowledges by saying that A "is right". What is the conversational function of "although, if you mean... you're right"? Through this remark, what B achieves is to take back or suspend her first assertion, with all the entailments that it carried, and not merely the existential entailment (as Haugh seems to think). The entire assertion that Bob Beamon long-jumped over 29 feet way back in 1968, although still true, becomes irrelevant at this point, and contributes nothing to the purpose of the conversation. So, one might say that this is a case of "cancelling" an entailment by taking back the whole assertion that carries it. But, of course, this is not what is normally understood by "cancelling" an implication. Instead, to cancel an implication is to cancel *only* the implication, not the assertion that carries it.

Moreover, what B does by using the conditional is not, strictly speaking, to cancel or retract her previous assertion, but only to call it into question by offering an alternative interpretation for A's utterance. Such calling into question of a content that one already asserted or implied is known as "suspension", which Sadock (1978, p. 292) and Levinson (1983, p. 115) distinguish from cancellation. In as much as the original interpretation has not been cancelled, but only suspended, the same is true of the entailment.[13]

I propose below a dialogue that contains what I take to be a better candidate for the strict cancellation of an entailment, and not mere suspension. Consider the following scenario: on her way to her office A takes the elevator closest to her

---

[13] In fact, Levinson notes that "entailments, being non-defeasible, cannot be suspended" (1983, p. 115)

office, the one everyone normally uses (call it E). A notices that E makes a strange noise, and, upon entering the office, she asks her colleagues:

(10)   A: Has anyone used the elevator this morning?
       B: I have used the elevator this morning. But I do not mean to imply that someone used the elevator closest to the office. In fact, this time I used the one which is further away.

With "the elevator" A refers to E. On one interpretation, the proposition literally expressed by B's utterance of the first sentence is that *B has used E this morning*. This proposition entails that *someone has used E this morning*, which is the affirmative answer to A's question. With her second sentence B cancels this interpretation, and makes it clear that she uses the verbal phrase "use the elevator" with its generic sense (similar, for instance, to "take the bus"). Does B manage to cancel (in the relevant sense) the entailment that *someone has used E this morning*? Again, the answer is no. With her second utterance, B cancels a certain possible, but unintended, interpretation of the utterance (i.e. that *B has used E this morning*) that carries the existential entailment (i.e. that *someone has used E this morning*). That is, what is cancelled is *not* merely an entailment of the proposition literally expressed, but a whole interpretation of the utterance. The intended interpretation does not carry the entailment that *someone has used E this morning*. And so, no entailment of what B actually said has been cancelled.

To sum up, in the previous section we have seen that CIs are calculable, but that this requirement does not rule out entailments. In this section we have seen that entailments do not seem to be cancellable (or, at least, that the alleged examples of cancellable entailments I have considered are not convincing). This suggests a negative answer to our initial question (i.e. entailments cannot be at the same time the content of a CI), but only if cancellability is a universal feature of CIs. We have seen that several objections have been raised to this last claim, although their success is a matter of ongoing controversy. That is, a negative answer to our initial question requires a deeper argument why CIs are *always* cancellable. If we lack such an argument we have no solid basis for a negative answer. We might equally well decide to accept entailments as candidates for CIs, and at the same time take cancellability to be a non-universal property of CIs.

Why should we even expect that all CIs be cancellable? One reason sometimes invoked in the literature is that pragmatically conveyed contents are expected to be cancellable, while contents that are semantically encoded are not. As Sadock (1978, p. 292) notes, "[c]ontradicting any part of the conventional meaning of an utterance [...] will amount to logical contradiction (in the case of

semantic content)." What is conventionally expressed, i.e. encoded by the linguistic meaning of the sentence, be it part of what is said, a semantic presupposition, or a conventional implicature, cannot be cancellable. This suggests that we need to turn to the last property on our list, (b) non-conventionality.[14]

5 Non-conventionality

Of the various criteria for CI that we have considered so far, calculability and non-conventionality are the two defining properties, according to Grice. Although he offers no official definition of CI in terms of sufficient conditions, he does give what might be taken to be an *approximation* to a definition when he writes: "I wish to represent a certain subclass of nonconventional implicatures, which I shall call *conversational* implicatures, as being essentially connected with certain general features of discourse" (1989, p. 26).[15] While calculability, as we have seen, does not exclude entailments from being the content of CIs, non-conventionality does. So, Grice's official (approximate) definition of CIs excludes entailments, on the basis of the last one of the eight criteria we have considered, non-conventionality. If we accept the Gricean definition, no proposition that is entailed can be at the same time the content of a CI.

This might be taken as the definitive answer to our question, if we want to be faithful to Grice's own approach to implicature. One could either accept this conclusion, which provides a negative answer to our initial question, or alternatively reconsider the very definition of CI. A reason to explore this latter alternative is that we have found no other reason to exclude entailments from the category of CIs apart from Grice's definition of CI. So, in what follows I consider the possibility of revising this definition, being aware that in doing this I am departing from Grice's own view of the matter.

---

14 An anonymous reviewer points out that this claim is false, given that there are examples of conventionally encoded implications that are cancellable, such as the pragmatic presuppositions introduced by "regret". Instead, s/he argues that it is the fact that entailments are part of the truth-conditional content of the utterance that makes them non-cancellable. If this observation is correct, then non-conventionality is not the deeper reason why entailments are non-cancellable. Still, the non-cancellability of entailments does not offer a good basis for either an affirmative or a negative answer to our initial question. One way or another, a discussion of non-conventionality is needed.

15 Sadock (1978, pp. 284–285) also writes that CIs are "by definition nonconventional".

Are there good reasons to take non-conventionality as part of the definition of implicature? There are at least two reasons why one might insist that conventionality needs to be part of the definition of CIs. I discuss them separately in what follows.

As Sadock (1978, p. 285) notes, one reason why calculability is not sufficient to characterize CI is that it overgenerates predictions of CI.[16] In particular, consider dead metaphors. These are expressions that had a use which was metaphorical at a certain time in the past, say before T. Before T the use is best explained as involving a CI that is calculable given the literal meaning (call it L1) of the expression. However, the use became lexicalised over time, and so it ceased to be metaphorical after T. Starting from T onward, the use is best explained by postulating a second lexical meaning for the expression (call it L2). For instance, as Reimer (1998, pp. 97–98) explains, the verb "to incense" was once used with the meaning *to make fragrant with incense* although it is now more commonly used to mean *to make very angry*. The newer meaning is not metaphorical at the present moment, although it was at a certain point, before it became lexicalized. But a pragmatic explanation of that use in terms of a CI is still available after T. Even after a metaphor dies and no speaker relies any longer on inferential processes in interpreting it, a Gricean derivation that fulfils the calculability requirement of the new meaning (L2) from the old one (L1) is still available. But this account of the use is no longer correct, so we need to rule it out. One way to achieve this is by introducing the requirement of non-conventionality of CIs in their definition. Thus, the new meaning is not a CI because it is conventionalized, and CIs are non-conventionalized.

Is this a good reason to maintain non-conventionality as part of the definition of CI? I think it is not. The reason is the following: when the metaphor dies, and the use is lexicalized, the word achieves a new literal meaning L2, apart from the old one, L1. The word is now ambiguous, having two meanings L1 and L2. When a competent speaker uses the word in the second sense she will deploy L2 and will not derive this sense indirectly from L1. There is no need for her to go through an indirect derivation of the meaning if she already has at her disposal a direct resource to express that meaning. The fact that there still is an alternative route to reach that meaning that involves a CI on top of L1 is irrelevant, because that is simply not the correct explanation of the use in this case. So, the requirement of

---

16 See also Wilson and Sperber (1986, pp. 107–108) for a similar point.

non-conventionality is not needed to stop the overgeneration of CIs for cases of lexicalized uses, because the latter simply does not occur.[17]

The second reason why non-conventionality is taken to be a defining feature of CI has the purpose of distinguishing CI from conventional implicatures. Grice introduces CIs as "a certain subclass of nonconventional implicatures" (1989, p. 26). The question about the conceptual relation between conventional and conversational implicatures requires a clear understanding of both concepts. While typical examples of conventional implicature involve 'therefore' (Grice 1989, p. 121), and 'but' (Grice 1989, p. 88), there is some disagreement on what exactly enters into the category, and on how exactly to define it. Potts (2005) includes in the category *syntactically* triggered conventional implicatures, such as relative subordinate clauses, as in the following example:

(11) Jill, who missed her flight from Madrid to Barcelona, took a train to get home.

Sadock (1978, p. 292) includes among conventional implicature the implication that factive verbs introduce that the proposition expressed by their sentential complement is true, such as the implication in the statement, "It's odd that dogs eat cheese." that *dogs eat cheese*. However, most authors take such implications to be *presuppositions*, instead of conventional implicatures (Beaver and Geurts 2011). On the other hand, Bach (1999, 2006, section 10) argues that not even the most popular examples of conventional implicatures are genuine instances of this class. Despite this multiplicity of positions, there is some agreement on the *definition* of the category.[18] Many authors, including Bach (1999, p. 331), Potts (2005, p. 11), Barker (2003, p. 3), and Horn (2008, p. 181), follow Grice (1989) in defining "conventional implicature" as implications that have two features: they are *non-truth-conditional* implications of an utterance of a sentence[19] (in the sense

---

17 Of course, this does not help too much in practice in order to *identify* the presence of a CI, and to find out whether we are before T or after T with regards to a certain use of an expression (and so whether one or the other explanation of the use is correct). In finding this out we might want to rely on intuitions about whether a meaning is conventionalized or not. As Grice (1989, p. 49) notes, "[w]e must of course give due (but not undue) weight to intuitions about the existence or nonexistence of putative senses of a word (how could we do without them?)".

18 However, Potts (2005, p. 5) notes that many semanticists use the term synonymously with 'presupposition', implicitly denying that there is a proper category to be defined.

19 The fact that they are irrelevant to the truth-conditions of the utterance is the semantic feature that distinguishes conventional implicatures from at-issue entailments and semantic presuppositions.

that their propositional content is not dependent upon, and does not contribute to, the proposition literally expressed) and *conventional* (i.e. lexically triggered, introduced by the linguistic meaning of the words uttered).

Other definitions, such as the ones in Leech (1983, p. 11) and in Levinson (1983, p. 127), explicitly add a third feature to the definition of conventional implicatures: *non-calculability*.[20] For instance, Levinson (1983, p. 127) writes: "Conventional implicatures are non-truth-conditional inferences that are *not* derived from superordinate pragmatic principles like the maxims, but are simply attached by convention to particular lexical items or expressions." We grasp the content that a conventional implicature introduces in virtue of our linguistic competence, and not by way of an inference to the best explanation of what the speaker means which draws on considerations concerning conversational cooperation. Moreover, a derivation *could not* give us the content of the conventional implicature. As Williamson (2009, p. 150) notes, "if there were not already a significant difference between the words 'Boche' and 'German', the norms of conversation would not generate any difference in implicature between 'Lessing was a Boche' and 'Lessing was a German'." If a hearer fails to grasp the linguistic meaning of a word such as 'Boche', she cannot recover the pejorative implication from the context. That is, conventional implicatures are not calculable from what is said, information concerning the context of utterance, and assumptions concerning the conversational maxims. But if conventional implicatures are non-calculable, and CIs are calculable by definition, we already have a criterion to distinguish them. And so, there is no need to introduce non-conventionality as a defining feature of CIs. Actually, as Grandy (1989, p. 518) pointed out, Grice himself suggests in a well-known passage that calculability is *sufficient* to mark off those implicatures that are conventional from those that are conversational. He writes:

> The presence of a conversational implicature must be capable of being worked out; for even if it can in fact be intuitively grasped, unless intuition is replaceable by an argument, the implicature (if present at all) will not count as a conversational implicature; it will be a conventional implicature.
>
> (Grice 1989, p. 31)

---

20 Bach (1999, p. 329) also notes that conventional implicatures are not calculable.

That is, an implicature that is not conversationally calculable is conventional. And so, the second reason given for introducing non-conventionality into the definition of CI is not good enough.[21]

In conclusion, I am departing slightly from Grice's official definition of CI in proposing to take non-conventionality out of it, and to define CIs simply as *calculable* implications.[22] This definition of CI does distinguish them from conventional implicatures, as the latter are not calculable. I propose to use the term "conversational implicature" as synonymous with "calculable implications", where "calculable" refers to a certain calculation that relies on assumptions concerning the conversational maxims. Conventional implica*tures* are distinguished from conversational ones because the former ones are not calculable. But some conventional implica*tions* of utterances of sentences could be conversational implicatures. This is the case with entailments.[23] As a result, the answer that I propose to our initial question is affirmative: an entailment can be the content of a conversational implicature. Given that what is definitory for the latter category is the calculability requirement, entailments are not eliminated from this category by default. Contrary to what Laurence Horn and other authors quoted in the first section of this paper maintain, being a conversational implicature and being an entailment of the semantic content are not really different categories within the same classification. Instead, according to my proposal, an implication has the status of *entailment* if it bears a certain logical relation to the semantic content of the utterance; and it has the status of a *conversational implicature* if it is conversationally calculable.

---

21 However, Neale (1992, p. 527) disagrees with Grandy on this interpretation of Grice (1989) and writes that a definition of CI that only includes calculability does not exclude conventional implicatures.

22 A reason why one might doubt that calculability is a sufficient condition to define CIs is that the disambiguation of an ambiguous utterance of a sentence, or the interpretation of the reference of an indexical, involves some sort of calculation as well. However, what one calculates in these cases is the semantic content of the utterance, not an implication of a certain kind. In the case of CIs one *relies* on the semantic content of the utterance in order to calculate an implication. And CIs are defined above as calculable *implications*.

23 Of course, a content cannot be at the same time an entailment and a conventional implicature, as the latter is both conventional *and* non-truth-conditional, while entailments are truth-conditional.

6 Concluding remarks

I have argued in this paper that an affirmative answer to the question whether entailments could figure as contents of CI is warranted. In particular, the two features of CI that could rule out entailments from the class of contents that could be conversationally implicated are (a) cancellability, and (c) non-conventionality. Entailments are non-cancellable, but this is a reason to conclude that they cannot be CIs only if cancellability is a universal property of CIs; alternatively, one might accept CIs that are entailed by what is said and, on this basis, reject the claim that all CIs are cancellable. I see no compelling reason to go one way or another. So, criterion (a) does not lead to a principled answer to our question. Turning to (b) non-conventionality, this is a defining feature of CIs, according to Grice's official definition. I explored the reasons given in the literature for taking non-conventionality to be a defining feature of CI, and argued that none of them are compelling: lexicalization of dead metaphors does not lead to overgeneration of CIs, and calculability is sufficient to distinguish conventional from conversational implicatures. I propose a definition of CI in terms of calculability only, which does not rule out entailments. So, the answer to our initial question is, eventually, a tentative 'yes'.

A notable consequence of this definition is that the argument against the cancellability test based on cases of CIs the content of which is entailed by what is said seems a good one after all. This means that CIs are not necessarily cancellable. However, that does not mean that the test is a bad one, inasmuch as non-entailed CIs are still expected to be cancellable.

References

Bach, Kent (1994): "Conversational Impliciture". In: *Mind and Language* 9(2), pp. 124–162.

Bach, Kent (1999): "The Myth of Conventional Implicature". In: *Linguistics and Philosophy* 22(4), pp. 327–366.

Bach, Kent (2006): "The Top 10 Misconceptions About Implicatures". In: Betty Birner/Gregory Ward (Eds.): *Drawing the Boundaries of Meaning: Neo-Gricean pragmatics and semantics in honor of Laurence R. Horn*. Amsterdam: John Benjamin, pp. 21–30.

Barker, Stephen (2003): "Truth and Conversational Implicature". In: *Mind*, pp. 112: 1–33.

Burton-Roberts, Noel (2010): "Cancellation and Intention". In: Belén Soria/Esther Romero (Eds.): *Explicit Communication. Robyn Carston's Pragmatics*. London: Palgrave Macmillan, pp. 138–155.

Blome-Tillmann, Michael (2008): "Conversational Implicature and the Cancellability Test". In: *Analysis* 68, pp. 156–160.

Blome-Tillmann, Michael (2013): "Conversational Implicatures (and How to Spot Them)". In: *Philosophy Compass* 8 (2), pp. 170–185.

Borge, Steffen (2009): "Conversational Implicature and Cancellability". In: *Acta Analytica*. 24, pp. 144–154.

Capone, Alessandro (2009): "Are Explicatures Cancellable? Towards a Theory of the Speaker's Intentionality". In: *Intercultural Pragmatics* 6, pp. 55–83.

Carston, Robyn (2002): Thoughts and Utterances. The Pragmatics and Explicit Communication. Oxford: Blackwell.

Colonna Dahlman, Roberta (2013): "Conversational Implicatures Are Still Cancellable". In: *Acta Analytica* 28, pp. 321–327.

Davis, Wayne A. (1998): *Implicature: Intention, Convention, and Principle in the Failure of Gricean Theory*. Cambridge: Cambridge University Press.

Grandy, Richard (1989): "On Grice on Language". In: *Journal of Philosophy* 86, pp. 514–525.

Grice, H. Paul (1981): "Presupposition and Conversational Implicature". In: Peter Cole (Ed.): *Radical Pragmatics*, New York: Academic Press, pp. 183–198.

Grice, H. Paul (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Haugh, Michael (2013): "Implicature, Inference and Cancellability". In: Capone Alessandro/Franco Lo Piparo/Marco Carapezza (Eds.): *Perspectives on Pragmatics and Philosophy. Perspectives in Pragmatics, Philosophy & Psychology*, vol 1. Cham: Springer, pp. 133–151.

Higashimori, Isao/Wilson, Deirdre (1996): "Questions on Relevance". In: *UCL Working Papers in Linguistics* 8, pp. 111–124.

Horn, Laurence (2008): "'I Love Me Some Him': The Landscape of Non-Argument Datives". In: Olivier Bonami/Patricia Cabredo Hofherr (eds.) *Empirical Issues Syntax and Semantics 7*, pp. 169–192. Available at: http://www.cssp.cnrs.fr/eiss7.

Horn, Laurence (2012): "Implying and Inferring". In: Allan Keith/Kasia M. Jaszczolt (Eds.): *The Cambridge Handbook of Pragmatics*. Cambridge: Cambridge University Press, pp. 69–86.

Kaplan, David (1989): "Demonstratives – An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals". In: Joseph Almog/John Perry/Howard K. Wettstein (eds.): *Themes from Kaplan*, Oxford: Oxford University Press, pp. 481–526.

Leech, Geoffrey N. (1983): *Principles of Pragmatics*. London/New York: Longman.

Levinson, Stephen C. (1983): *Pragmatics*. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Levinson, Stephen C. (2000): *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Moeschler, Jacques (2012): "Conversational and Conventional Implicatures". In: Hans-Jörg Schmid (Ed.): *Cognitive Pragmatics*, Berlin: De Gruyter Mouton, pp. 405–536.

Neale, Stephen (1992): "Paul Grice and the Philosophy of Language". In: *Linguistics and Philosophy* 15, pp. 509–59.

Potts, Christopher (2005): *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Reimer, Marga (1998): "Donnellan's Distinction/Kripke's Test". In: *Analysis*. 58(2), pp. 89–100.

Sadock, Jerrold (1978): "On Testing for Conversational Implicature". In: Peter Cole (Ed.): *Syntax and Semantics*, Vol. 9: Pragmatics, New York: Academic Press, pp. 281–298.

Sperber, Dan/Deirdre Wilson (1995): *Relevance. Communication and Cognition* (2nd edn). Oxford: Blackwell.

Vicente, Begoña (1998): " Against Blurring the Explicit/Implicit Distinction". In: *Revista Alicantina de Estudios Ingleses* 11, pp. 241–258.

Weiner, Matthew (2006): "Are All Conversational Implicatures Cancellable?" In: *Analysis* 66, pp. 127–130.

Wilson, Deirdre/Sperber, Dan (1981): "On Grice's Theory of Conversation". In: Paul Werth (Ed.): *Conversation and Discourse*. London: Croom Helm, pp. 155–178.

Wilson, Deirdre/Sperber, Dan (1986): "Inference and Implicature". In: Charles Travis (Ed.): *Meaning and Interpretation*. Oxford: Basil Blackwell, pp. 45–75.

Williamson, Timothy (2009): "Reference, Inference, and the Semantics of Pejoratives". In: Joseph Almog/Paolo Leonardi (Eds.): *The Philosophy of David Kaplan*. Oxford: Oxford University Press, pp. 137–159.

Åkerman, Jonas (2015): "Infelicitous Cancellation: The Explicit Cancellability Test for Conversational Implicature Revisited". In: *Australasian Journal of Philosophy* 93 (3), pp. 1–10.

Pedro Santos

Trying to Make Sense of Embedded Conversational Implicatures

Abstract: Attempts at accounting for so-called embedded conversational implicatures (CIs) come up against the basic problem of making sense of kinds of content that, despite being derived from *bona fide* conversational implicatures, seem to be part of the literal meaning of utterances. The problem resists the conventionalist way out (championed, among others, by Chierchia and Levinson), which construes generalized CIs as not meriting the status of CIs proper, this being one way of dealing with the puzzle. One decisive drawback of this strategy is that particularized CIs, which could not credibly be conventionalized, display the same kind of puzzling embedding behavior as generalized CIs. In this paper I discuss Mandy Simons's and François Recanati's approaches to the conundrum. Simons assigns to embedded CIs the status of *bona fide* Gricean inferences, building on the distinction between embedded pragmatic *effects* and embedded pragmatic *computations* and claiming that the puzzle is solved by the realization that only the latter phenomenon is really problematic and that embedded CIs are to be analysed as cases of the former rather than as cases of the latter. Recanati has (arguably justifiable) qualms about this sort of approach and attempts to account for the phenomenon by dismissing (somewhat in the vein of the conventionalists) the very notion of an embedded CI as intrinsically incoherent. He purports to do this on the basis of a globality criterion on Gricean inferences that he argues cases of alleged embedded CIs only partially fulfil, so that they count as cases of Gricean inferences *latu sensu* (as, for instance, are cases of what he calls modulation) but surely not as genuine CIs. I argue that Recanati's take on the problem cannot be successful, as globality is a faulty criterion as far as distinguishing cases of (embeddable) modulation from CIs proper is concerned.

Keywords: pragmatics, conversational implicature, embedded pragmatic effects

1 Preliminary observations

The notion that pragmatics intrudes into what some authors would call semantic content, but could perhaps also be called literal truth-conditional content,[1] is now widespread and much less contentious than it used to be. Recent (and not quite so recent) work in the philosophy of language provides a fair amount of evidence for such an anti-Gricean picture of the relationship between literal and pragmatic meaning. The contextualist literature in particular offers a fair number of cases of (to use Recanati's terminology) primary pragmatic effects, i.e. pragma-tically derived content with truth-conditional import. Indexicals, gradable adjectives, and quantifiers, for instance, arguably provide examples of such an intrusion of pragmatics into literalness. The basic intuition those accounts purport to elaborate on is that on many occasions the propositions literally expressed by utterances and on the basis of which they are to be assigned a truth-value go beyond the semantics of the uttered sentence. More contentiously, contextualists have emphasized the theoretical import of (alleged) evidence for the existence of what Recanati calls modulation. These are non-linguistically mandated, optional pragmatic processes that nevertheless, it is argued, fall short of meriting the status of conversational implicatures (henceforth, CIs), accordingly being included by Recanati in the category of "primary pragmatic processes". *Pace* recent attempts by "minimalist" authors like Emma Borg[2] to defuse the theoretical import of this kind of evidence, these cases seem to show that a lot of what counts as literal truth-conditional content can reasonably be accounted for in part in terms of pragmatics – i.e. in terms of the communicative intentions of speakers and, correspondingly, in terms of decoding strategies on the part of interpreters that go well beyond their semantic competence, largely drawing on their mastery of Gricean maxims.

---

1 Elaborating on this terminological shift would entail discussing the criteria for drawing a borderline between semantics and pragmatics, which I will not attempt to do here. In any event, it is worth pointing out that I will be assuming throughout this paper a broadly "Recanatian" stance on the issue. In particular, my references to an utterance's "literal" content are not to be understood as merely singling out its linguistically encoded meaning, covering instead the output of what Recanati often calls "primary pragmatic processes", yielding pragmatically derived propositions that nevertheless fall short of being conversationally implicated. As will become clear below, this will be taken to include the kind of pragmatically enriched material often referred to as "embedded conversational implicatures".

2 Cf. for instance Borg (2012).

Further evidence for pragmatic intrusion into the realm of literal meaning comes from so-called embedded CIs, i.e. pragmatically derived content that occurs under the scope of some operator and therefore would not normally count as conversationally implicated, but which otherwise displays the features of canonic CIs. In this paper I will discuss the question of whether the corresponding pragmatic effects should actually be assigned the status of CIs proper. I will essentially adopt Recanati's negative answer to this question, building on his claims[3] i) that CIs are kinds of inference whose basic premise is that an assertion has been made, and ii) that the inferences generating the embedded pragmatic effects often referred to as embedded CIs do not fulfil that condition. In the last section I will also argue, nevertheless, that his globality criterion for setting genuine CIs apart from other kinds of pragmatic inferences is faulty, so that a broadly Recanatian take on the phenomenon (which I generally endorse) has on the face of it no alternative than to resort to his infamous "Availability Criterion".

Typical cases of (so-called) embedded CIs are the following:

(1) If she knows some ministers, that might not be enough. She must know the Minister of Home Affairs.
(... some but not all ministers)
(2) She either saw him and shouted or the other way round.
(... and shouted after seeing him)

These are embedded pragmatic effects with a peculiar feature: they would normally count as genuine CIs when unembedded; they can therefore be seen as a particularly striking kind of evidence of pragmatic intrusion into literal content. They are also particularly problematic kinds of such intrusion. In the literature, cases of embedded CIs have often been dismissed as semantic (or pragmatic, or semantic and pragmatic) fictions. In fact, on the face of it, condoning the possibility of CIs that occur in an embedded environment (typically, under the scope of some operator) presupposes the acceptance of the undoubtedly problematic (and arguably incoherent) notion that there are genuine CIs which contribute to the literal meaning of utterances. In a sense, then, intuition has it that CIs cannot embed. And this particular intuition is hard to evade.

L.J. Cohen pioneeringly noted the existence of what came to be called embedded CIs (Cohen 1971). As it happens, he was mainly interested in arguing against the Gricean pragmatic account of indicative conditionals on the basis of examples like *If the old king has died of a heart attack and a republic has been declared, then*

---

3 Cf., among others, his (2017).

Tom will be quite content. These display what could be argued to be an embedded CI, implicating that the temporal ordering that relates the states of affairs corresponds to the order by which they are being referred to in the antecedent, building on the maxim of Manner, along the same lines as the one in (2) (only in this case in the antecedent of a conditional rather than in a disjunct). Cohen's own argument was that, since the temporal ordering is part of the literal content of the antecedent of conditionals, indicative conditionals cannot be truth-functional if the conjunction is, and hence Grice's views on these structures are inherently contradictory. This particular kind of skepticism need not, of course, affect more than Grice's proposal for a truth-functional treatment of indicative conditionals. But Cohen's example does seem to show that there are embedded CIs – which, as suggested above, is intuitively problematic.[4]

Not only is the notion of an embedded CI intuitively problematic, but it can also arguably be *established* that CIs cannot embed. The argument for the non-embeddabilty of CIs runs roughly as follows. Since Grice's pioneering discussion, conversational implicatures (CIs) are assumed to arise from inferential processes based on two elements:

  i. The fact that some speaker S uttered some sentence s
 ii. The fact that, in uttering s, S is committed to the Cooperative Principle (CP) and the conversational maxims, which in turn legitimatizes the assumption that his/her utterance complies with them

The general argument builds on the assumption that CIs are (most typically) derived from "what is said", i.e. from some asserted propositional content – hence, from a speaker's act of assertion. One arguably unavoidable consequence of this is what Recanati calls *globality*: CIs must be global, i.e. derived from the utterance of whole sentences. This, in turn, is taken to entail that no CIs can be derived from non-asserted, sub-sentential material, as this would make the inference "local" rather than "global". Here is the argument in schematic form:[5]

No Embeded CIs (NECI)
1. Conversational implicatures are pragmatic consequences of an act of saying something.

---

4 Other discussions in the literature of embedded pragmatic processes or inferences (notably by Relevance Theorists, cf. Sperber and Wilson (1986) and Carston (2002), as well as by other contextualist-minded authors like Recanati) are, in a way, sequels to Cohen's pioneering remark.
5 Recanati, who presents the argument in his (2017), attributes it to Anscombre and Ducrot (1983) in a previous formulation of it, in his (2003).

2. An act of saying something can be performed only by means of a complete utterance, not by means of an unasserted clause such as a disjunct or the antecedent of a conditional.

3. Hence, no CI can be generated at the sub-locutionary level, i.e. at the level of an unasserted clause such as a disjunct or the antecedent of a conditional.

4. To say that a CI falls within the scope of a logical operator is to say that it is generated at the sub-locutionary level (i.e. at the level of the clause on which some logical operator operates).

5. Therefore, no CI can fall within the scope of a logical operator.

This argument spells out what Simons calls the "calculation problem" concerning so-called embedded implicatures. How can it be that embedded material (e.g. material under the scope of an operator) is the input for a CI, given that CIs are generated out of asserted material[6] and embedded material is by definition not asserted? In other words, if the argument is sound, the notion of an embedded CI is shown to be incoherent: since CIs are by definition "global" and since "globality" entails non-embeddability, there cannot be any embedded CIs. And, given that there are no embedded CIs, the evidence that allegedly vouches for their existence has to be accounted for in some other way.

   A tendency present in much of the literature on "embedded CIs" is the conventionalist school of thought, championed by, among others, G. Chierchia and S. Levinson. Underlying its various (and theoretically diverse, as symbolized by these two authors[7]) implementations is the claim that scalar inferences can generally embed, and that such behaviour is inconsistent with assigning them the status of genuine CIs. The approach crucially builds on the fact that those inferences are canonical examples of what Grice calls generalized conversational implicatures (GCIs), i.e. CIs which (as opposed to *particularized* CIs) are generally associated with the way a certain expression is taken to comply with the Cooperative Principle. Thus, in a GCI, *p* implicates *q* in the sense that, in most (or in the typical, or unexceptional) contexts of utterance, a speaker asserting *p* complies with the CP only if she also means *q*, so that we can infer *q* from her asserting *p* without any need to rely on any specific contextual information; the inference goes through by default, so to speak. To quote Green 1998: "If an assertion of a sentence S conveys the *implicatum* that p with nearly universal regularity, then when S is embedded the content that is usually

---

6 As is often the case, I am taking assertion to be the model case for the generation of a CI, although assertion is here to be interpreted somewhat metonymically as standing for any kind of illocutionary act that could generate a pragmatic inference.

7 See Chierchia (2004) and Levinson (2000), among others.

understood to be embedded for semantic purposes is the proposition S & p" (Green 1998, p.72). Thus the implicatures in (1) and (2) would, if unembedded, be obvious examples of a GCI.[8] Hence the cogency of taking them to be semi-conventionalized implicatures or "presumptive meanings" a la Levinson (2000). Those that can be seen in that way can also occur in embedded contexts, but that fact in itself is taken to be evidence that they are not genuine CIs.[9]

One potentially unsurmountable drawback of the conventionalist approach for dealing with (alleged) embedded CIs is that particularized CIs, for the treatment of which the conventionalist strategy is clearly not available, can also embed. The following examples, involving disjunctive and the conditional operators, are borrowed from Simons (2017):

(3) A: What are you going to do for your mother's birthday?
    B: Either I'll buy flowers, or I'll take her to a nice restaurant.
    (*buy flowers* understood to mean "buy flowers as a birthday present to my mother")
(4) A: How will you get to Boston?
    B: Either I'll rent a car or I'll fly.
    (*rent a car* understood to mean "rent a car for the purpose of driving it to Boston")
(5) A: What's making noise up in the attic?
    B: I'm not sure. But if there's a nest up there, we're going to have a big mess to clean up.
    (*a nest* understood to mean "an occupied nest")
(6) [Making plans to get together for an evening]
    If you cook dinner, I'll bring dessert.
    (*cook dinner* understood to mean "cook dinner for us both")

In (3)/(4), for each disjunction to be conversationally acceptable, both disjuncts have to be possible answers to the respective questions. But for that to be the case,

---

8 I will not be discussing the merits of these proposals, or the cogency of the notion of a GCI, or the fuzziness of the line separating them from those CIs Grice classifies as particularized, as those issues have no direct import on this paper's argument.
9 Geurts (2009) shows that much of the evidence about the embeddedness of scalar inferences can be explained (away) without rejecting their CI status. In any event, if conventionalism and/or the way out suggested by Geurts for some of these cases is correct, the notion of a CI as non-embeddable by definition is preserved: those scalar inferences that embed *and* fall out of reach of Geurts's account are thereby shown not to be genuine CIs.

in each case the first disjunct must be enriched as illustrated, in view of the question being answered and given general considerations of relevance. In the conditionals (5)/(6), if the antecedent is given its literal interpretation, the resulting conditional makes a claim which the speaker is unlikely to intend to commit to. In (5), the presence of a possibly unused nest would not result in a mess; it is the birds' occupying the nest that would be the problem; in (6), the speaker is unlikely to offer to bring dessert simply as a reward for you cooking dinner for yourself. It is noticeable that in all four cases the pragmatic inferences would count not as generalized but as particularized CIs (i.e. heavily context-sensitive inferences, only derivable in view of certain contextual cues), if they are to be seen as CIs at all. In other words, we need a non-conventionalist account of what is going on in these cases, a way of making sense of these apparent cases of embedded pragmatic inference.

2 Simons and the continuum

Mandy Simons's approach to the phenomenon (see Simons 2014, 2017) rejects one basic tenet of conventionalism: she defends the idea that a Gricean kind of approach is powerful enough to tackle embedded pragmatic effects and, specifically, the idea that all cases of embedded pragmatic inferences which, when unembedded, would count as CIs, are akin to CIs when embedded, too. She therefore also rejects NECI, going as far as stating that these "embedded" inferences "can be seen as continuous with ordinary, utterance-level, conversational implicatures".[10] Simons is aware that construing cases of embedding as CI-like inferences is not free from difficulties, especially when one tries to spell out the details of such an approach. Nevertheless, she claims this should be the preferred theoretical option.[11]

From a pre-theoretical point of view, so-called embedded CIs are, of course, particularly good candidates for a Gricean type of analysis. First and foremost, as already mentioned, they stand out among pragmatic inferences as being distinct from those associated with indexical resolution or reference assignment, in that

---

10 Simons (2017, p. 468).

11 In her own words: "There is no reason why, as theorists, we have to take Grice's views as an unseparable whole. The fundamental ideas underlying the notion of implicature are clearly separable from Grice's expressed views about the nature of "what is said" and about the semantic content of particular items. (…) One can maintain a broadly Gricean outlook while modifying the details of the account of how implicatures are generated" (Simons 2010, p. 142).

they are not linguistically mandated, being based on purely pragmatic considerations. As pragmatic inferences, they are abductive paths to the most likely or most reasonable speaker's communicative intentions; but, furthermore, they essentially run, as much as CIs, on non-linguistic cues. Secondly, and more obviously, an intuitive motivation for Simon's CI-based account of the phenomenon is the fact that the inferences in question seem to correspond to the very same Gricean "calculus", i.e. they seem to have the same premises and the same conclusion (the derived pragmatic content) and they seem to appeal to the same conversational maxims as their unembedded, and *bona fide* CI, counterparts – hence the intuitive appeal of the very designation "embedded CIs".

But Simons's claim that there is no obvious or clear-cut difference between pragmatic effects such as those in (3)-(6) and their *unembedded* counterparts is, as already suggested, clearly problematic. A basic worry can perhaps be spelled out as follows. Intuitively, there does seem to be a fundamental similarity, indeed a continuum, between the cases where a Gricean inference operates on whole utterances and those where it operates on embedded material; but since in the former case the inference generates *implicated* content and in the latter the corresponding pragmatic effects could not possibly count as CIs, due precisely to their embeddability (at least if we take NECI seriously), the very thesis that there is such a continuum seems, under closer examination, rather objectionable. On the other hand, to Simons's credit, it must be acknowledged that this alleged conceptual impossibility does fly in the face of the basic intuitions supporting her claim. In other words, how can (what is seemingly) one and the same pragmatic effect both be a CI and not be a CI, depending on whether the pragmatically enriched material is embedded? If both types of cases display the same pragmatic effect, based on the same Gricean calculus, it would seem reasonable to say that if one is a CI, then so must the other one be. If they are not quite the same kind of inference, what exactly could be the difference? The dilemma is then clear: the claim that there is a continuum between the so-called embedded CIs and their unembedded counterparts and the claim there is no such continuum seem equally objectionable. Let us call this the continuum problem.

Simons (2017, p. 468) is careful to add that it is not her goal "to argue that embedded pragmatic effects are conversational implicatures in the strict Gricean sense", specifying that those pragmatic effects could reasonably be classified as conversational implicitures (following Bach 1994) and that her real goal is to argue that "there is more to Gricean pragmatics than implicature". Simons's point seems to have some leverage and indeed to be somewhat consensual nowadays even outside of its initial contextualist boundaries, as is clear from this paper's

initial remarks. The Gricean notion of an inference that builds on a general cooperativity commitment on the part of speakers to generate some enrichment of linguistic meaning is arguably powerful enough to account for different types of cases: the non-implicature, pragmatic inferences associated with disambiguation, reference fixing, and indeed with those pragmatic effects Recanati usually describes as cases of modulation, i.e. those components of meaning which, despite not being linguistically encoded, are part of literally asserted content (or "what is said", in the widely used Gricean phrase or, in Relevance Theory jargon, "explicature"). In other words, her claim that there is a continuum running from the embedded pragmatic effects she is concerned with to the uncontentious CIs those effects would be when unembedded could be read as the mere statement that Gricean inferences *latu sensu* are arguably at work in generating pragmatic effects of various kinds. In other words, it could be taken to be consistent with the preservation of the distinction between literal content and conversationally implicated content (i.e. the result of Gricean inferences *strictu sensu*), which Simons clearly does not want to question. But even if Simons's thesis is interpreted in such a noncommittal way, no hint of a solution to the continuum problem is forthcoming. So Simons's overall goal of showing that a Gricean approach is well equipped to account for embedded CIs does seem to be somewhat lacking in its conceptual credentials.

A possible reply to these worries is provided by Simons's distinction between *embedded pragmatic effects* and *embedded pragmatic computations*. The former concern the embeddedness of the output of pragmatic inferences, whereas the latter concern the embeddedness of their input. Here is how the distinction is articulated in Simons (2017, p. 470):

> *Embedded pragmatic effects*: Cases where the propositional content which falls under the scope of a linguistic operator (at some stage of interpretation) includes content which is the output of pragmatic inference.

> *Embedded pragmatic computation*: Cases where a pragmatic computation is carried out using as premises (at least) (a) the content of an unasserted embedded clause; or the observation that the speaker has expressed this content; or the observation that the speaker has expressed this content using a particular linguistic form *and* (b) general pragmatic principles.

The distinction is important for Simons's general take on so-called embedded CIs, since she intends to show that they are pragmatic inferences which generate embedded effects, but which, crucially, do not rely on embedded premises and therefore do not fall under the category of embedded pragmatic *computations*. In other words, she purports to argue for the coherence of the notion of a Gricean, global

inference which pragmatically enriches embedded material, but which does not run on local, unasserted premises. By this she means that embedded CIs do not involve consideration of the conversational appropriateness of the content occurring under a given operator. Rather, considerations of conversational appropriateness concern the utterance of the entire sentence of which the embedded material is a part, although the enrichment derived from such considerations only affects that embedded material. (In this rather indirect sense, and in this sense alone, could the embedded material be said to provide the input to the inference.) Simons also contends that even if not asserted, or intuitively not seen as asserted, the linguistically encoded material making up the uttered sentence is still liable to a Gricean-style "rational reconstruction". In other words, even if speakers do not actually assert what would be the propositional content expressed by the unenriched uttered sentence, the theorist can still acknowledge its role in an inference process akin to the generation of a CI.[12]

The following would be such a reconstruction of the inference leading to the enrichment of first disjunct in (3) from Simons's own description of it.[13] Let us recall that in (3) A asks a question and B replies with a disjunctive sentence. One of the disjuncts does not constitute, if interpreted literally, an answer to the question asked. Hence the pragmatic reasonableness of the enrichment. Let us focus on the first disjunct of (3) in an unembedded environment, as in (7):

(7) A: What will you give your mother for her birthday?
    B: I'll buy flowers.
    (... I'll buy flowers and give them to her)

Arguably, the enrichment can be accounted for in straightforward Gricean terms, according to the following sequence of steps:

  i. Speaker *A* asks a question. Speaker *B* is assumed to be a cooperative interlocutor, so that her response is a genuine answer to the question (there is no indication that her reply is intended to opt out of the dialogue)

  ii. *A* presumably knows that simply buying flowers is not something conventionally acceptable as a celebration of somebody's birthday, as opposed to *giving* flowers. *A* further recognizes that it is reasonable for *B* to expect her to infer that if *B* buys flowers in connection with her mother's birthday; it can

---

12 See, for ex. Simons (2017, p. 475).
13 Simons (2017, p. 474-76).

therefore be inferred that *B* intends to convey that she will buy flowers to give to her mother.

This Gricean type of inference naturally applies to the embedded version (3):

i. Speaker *A* asks a question. Speaker *B* is assumed to be a cooperative interlocutor, so that her response is a genuine answer to the question (there is no indication that her reply is intended to opt out of the dialogue)

ii. A disjunctive sentence given in answer to a question is felicitous only if *each disjunct* is interpretable as an answer to the question (a typically Gricean felicitousness constraint on disjunctions)

iii. *A* presumably knows that simply buying flowers is not something conventionally acceptable as a celebration of somebody's birthday, as opposed to *giving* flowers. *A* further recognizes that it is reasonable for *B* to expect her to infer that if *B* buys flowers in connection with her mother's birthday; it can therefore be inferred that, by means of the first disjunct of (3), B intends to convey that she will buy flowers to give to her mother.

This can be framed in a more canonically Gricean reconstruction of the reasoning that leads to the enrichment in question:

1. The content of the disjunction is compositionally determined.
2. Then it must be considered whether the disjunctive utterance is pragmatically adequate.
3. The first disjunct fails to provide an answer to the question asked.
4. Thus, by the constraint on disjunctions, so does the whole disjunction.
5. Therefore, the whole utterance, considered at the level of linguistically encoded content, is pragmatically inadequate.
6. But it is assumed that, as a participant in the interchange, the speaker is being cooperative.
7. So she must intend to convey something other than what her utterance literally conveys.
8. In particular, the root of the impropriety being the first disjunct, it is that disjunct that must be enriched in order for the whole utterance to comply with Cooperativity (specifically, with the maxim of Relevance).

Summarizing: the speaker has asserted *p or q*; but *p* is not in itself a cooperative answer to the question, which makes the utterance as a whole uncooperative (in particular, irrelevant). So, *p* cannot be the intended content of the first disjunct; a contextually relevant (minimal) enrichment of it must generate that content.

It should be obvious that the generation of the enriched meaning in (3) is tantamount to a straightforwardly Gricean CI calculus. In essence, one and the same inference seems to be at work in the generation of both the embedded and the unembedded effects. The difference is that in the latter case it determines only the content *of the disjunct*, which is why the output is an embedded pragmatic effect. But, crucially, although that output is the enrichment of a disjunct, it was not the disjunct in and of itself that triggered the inference. For the inference to run, it was indeed necessary for the interpreter to recognize that *the content of a particular disjunct* failed to satisfy the pragmatic appropriateness condition for disjuncts. Nevertheless, no Gricean reasoning taking that content as input (as asserted content in a standard CI) was invoked. Rather, the inference builds on the role of that disjunct *as a disjunct*, i.e. as a constituent of the whole disjunction – it crucially builds on the fact that the unenriched disjunct is uncooperative and therefore inappropriate *within the disjunction*. The Gricean inference was ultimately triggered by the observation that the disjunction *as a whole* failed to satisfy the requirement of Relevance (in this case, relevance with respect to the question asked). Without the premise that the speaker's whole utterance must in some way comply with the Cooperative Principle, the inference leading to the (embedded) pragmatic enrichment would not go through. Thus in both the embedded and the unembedded cases the generation of a pragmatic effect was based on cooperativity considerations regarding the utterance, globally considered. In neither case was the pragmatic enrichment generated at the level of a clause over which a logical operator has scope, even though in the embedded case that enrichment affects only non-asserted material. As in a proper CI, in this latter case the pragmatic enrichment seems very close to being the upshot of an assertive speech act.

If Simon's analysis is correct, Recanati's argument for the impossibility of embedded CIs must be somewhat faulty. In particular, premise 4 of NECI does not seem to be true. Embedded material (material under the scope of an operator such as the disjunction) can be enriched by a pragmatic, utterance-level inference without itself being the input for it; it is not the case that the inference is "generated at sub-locutionary level". In stating that the occurrence of a pragmatic effect applying to material under the scope of an operator entails that it is also *generated* under its scope, premise 4 unwarrantedly conflates embedded pragmatic *effects* with embedded pragmatic *computations*. Thus there is arguably no "calculation problem" to speak of. In no way does the above account of embedded pragmatic effects as illustrated by the disjunctive case assume that those effects are derived

from non-asserted material: the embeddedness concerns the output of the inference, not its input; the pragmatic effect, not the pragmatic computation. The pragmatic computation, following the Gricean canon, is global rather than local.

The overall conclusion is that Simons's "continuous" claim seems to be on the right track. The possibility of a Gricean reconstruction of the embedded effects, as shown in the disjunctive example, argues for the basic indistinctiveness between the pragmatic inferences generating those effects and those generating CIs. In both cases we have utterance-level inferences, easily accommodated within a Gricean account. The intuition of a continuum is vindicated, and the notion that there must be a fundamental difference between pragmatic inferences in cases like (3) and in cases like (7) is shown to be unwarranted. The continuum problem is thus arguably dissipated.

3 Globality

Despite their initial cogency, Simons's associated claims that premise 4 is false and that there is no "calculation problem" are perhaps too hasty. Recanati (2017) offers some useful criticism. His argument runs by and large as follows. Bearing in mind that, in a CI, we have to account not only for the way pragmatic effects are triggered but also for the generation of their effective content, we have to accept a crucial distinction Simons has not done sufficient justice to; namely, the distinction between *triggering* and *generating* pragmatic inferences.[14] Thus, in a pragmatic inference there are typically two sub-inferences:

i. The first leads to the conclusion that the speaker means something other (or, in some cases, something further[15]) than her literal utterance meaning (the triggering inference).

ii. From the conclusion that something else is (also) meant, the interpreter then performs a second, abductive inference, attempting to determine what the

---

14 To be fair, although not exploiting the distinction in the way suggested by Recanati, Simons acknowledges it, referring in particular to Recanati's "generating inference" as the "interpretative step" in a Gricean inference, e.g. in her (2017, p. 472).

15 Despite the fact that Recanati's actual examples mostly feature cases of substitution rather than supplementation (i.e. those cases often seen as resulting from the "exploitation" of a Gricean maxim as opposed to cases of non-exploitation), the dichotomy triggering/generating also applies to the latter.

speaker is most likely to have meant, given her commitment to the Coopera-
tive Principle (the generating inference)

Triggering inferences arise, then, from the fact that, under the presumption of co-
operativity, what the speaker means must differ from the linguistically encoded
content of the relevant utterance. Hence their *triggering* nature: they trigger an-
other inference, the one which leads to identification of the speaker's intended
meaning. That second inference (the *generating* inference) generates the overall
correct interpretation for the utterance, which either supplements or replaces its
linguistically encoded content. Recanati claims that in the case of genuine con-
versational implicatures the generating inference is *global*, since one of its crucial
elements is the premise that the speaker has actually made an assertion. By con-
trast, in cases of what he calls "pragmatic modulation", the generating mecha-
nism which is needed to yield the correct pragmatic interpretation cannot be a
global inference, since it does not include such a premise: these are cases where
the generation of a pragmatic effect does not stem from the actual assertion of the
pragmatically unenriched content.[16]

The same goes, Recanati contends, for cases of embedding as illustrated in
(3)–(6). According to Recanati, then, in these cases the premise that the speaker
has made an assertion is part of the triggering inference, which in that sense
counts as global. As in the cases of modulation mentioned above, from the fact
that an assertion has been made (and in the absence of any cues as to Coopera-
tiveness being flouted) speakers assume that the Gricean maxims are being ob-
served, which in turn warrants the conclusion that the linguistically encoded con-
tent of the utterance needs to be supplemented or replaced. But it is not as if the
content of that assertion is being taken as input for the *derivation* of a specific
pragmatic effect. Rather, it acts a trigger for the attempt to find some alternative,
non-linguistically encoded content that would enable the assertion to count as
cooperative. In other words, the fact that the linguistically encoded material in
the uttered sentence was asserted is not among the premises of the subsequent
generating inference. In cases like (3), for example, arguably no assertion is made
to the effect that one way of commemorating the speaker's mother's birthday is

---

16 One paradigmatic example would be metonymy, as in "The ham sandwich is getting restless"
uttered by a waiter as a means of referring to the person who ordered the ham sandwich (cf.
Recanati 2017, p. 495). Recanati's point is that in the generation of this kind of pragmatic effect,
no assertion of the linguistically encoded (absurd) proposition need to have taken place. In other
words, although the triggering inference arguably builds on the assumption that an assertion
has been made (one whose linguistically encoded content must be supplemented or replaced),
the same is not true of the generating inference.

merely to buy flowers; nor in (5) is an assertion made to the effect that the presence of just any nest in the roof is a sufficient condition for a mess to occur. No such propositional contents are being put forward for truth-evaluation. The respective pragmatic effects on the embedded material correspond to what, according to cooperativity concerns, would be the most reasonable interpretation of *that material*, regardless of what the linguistically encoded content of the whole utterance is. Thus in these cases generating inferences are strikingly distinct from those of genuine CIs: they are local rather than global.[17]

This is indeed what Recanati takes to be the essential difference between Gricean inferences *latu sensu* (of which the so-called embedded CIs are examples) and Gricean inferences *strictu sensu* (i.e. CIs proper): in the former case the generating inference is local, whereas in the latter it is global. Only the triggering inference is global in both cases. In other words, under this analysis embedded pragmatic processes are only partially global inferences, i.e. they are only global regarding the triggering sub-inference; they are not fully global inferences. This is their hallmark, and this is their basic distinctive feature with respect to CIs proper. In other words, those pragmatic inferences affecting material under the scope of an operator are not global enough, so to speak, to count as genuine CIs. Each of the corresponding inferences would amount to a CI if it occurred unembedded; but their embeddedness means that the linguistically encoded content that is their basic premise has actually not been asserted and therefore that no implicature calculus proper can have taken place.

Recanati's argument comes with a theoretical bonus. If he is right in claiming that pragmatic effects embed if and only if the corresponding generating inferences are local (i.e. if and only if those inferences are at best partially global), then he has provided a satisfactory criterion for distinguishing what he calls primary pragmatic processes (i.e. pragmatic processes which, like modulation, generate literal rather than implicated content) from secondary ones (i.e. CIs proper) and, simultaneously, with an account of why the former, but not the latter, embed. Full-fledged globality would then be the feature that singles out CIs, which do not embed, from among all other kinds of pragmatic inferences, which do.

---

17 Following Recanati's characterization, a straightforward way of making sense of the notion of globality is to say that a global (pragmatic) inference requires the assertion of a propositional content. But, unlike genuine CIs, in cases of embedded pragmatic effects the corresponding inference is not derived from an actual speech act of assertion, although one of its sub-inferences (the triggering one) assumes that some assertion has been made. The apparent contradiction can arguably be defused if we take the Gricean inferences which one would unwittingly describe as embedded CIs to be pragmatic inferences which, although global, are only partially so (since one of their sub-inferences is local).

The thesis that full-fledged globality is the basic criterion for setting CIs apart from all other kinds of pragmatic inferences has an important corollary: contrary to what Simons claims, premise 4 of NECI is not really faulty, and NECI, as a whole, is to be preserved. In fact, Simons's point that the embedded material is not input to the computation falls short of showing that there is no incoherence in the notion of an embedded CI. Granting that genuine CIs are characterized by full-fledged globality, the partial globality displayed by embedded pragmatic effects actually *prevents* them from falling under the category of CI; the embedded pragmatic effect is not generated by asserted content, as required by the notion of a CI. The incoherence of the notion of an embedded CI is sustained. So Simons's claim that a Gricean approach in terms of an implicature calculus is well-equipped to tackle embedded pragmatic effects seems to have little leverage after all.

Recanati's refusal of the status of CIs to embedded pragmatic effects provides one other nice corollary. Since such effects are based on pragmatic inferences that are only partially global and CIs are based on full-fledged global ones and are therefore totally different kinds of inference, the conceptual conundrum previously described as the continuum problem is thus defused. The fact that one and the same content can be the effect of similar pragmatic processes operating both on embedded and on unembedded material unduly gives the impression of a continuum; but Recanati seems to be right in claiming that the two kinds of inference are essentially diverse. Their similarity, although striking at first sight, is on reflection quite marginal.

4 A problem with the globality criterion

Unfortunately, Recanati's strategy for dealing with embedded pragmatic effects, and in particular with so-called embedded CIs, is not bulletproof. In this final section I will briefly discuss a counter-example to his globality criterion for identifying, from among all other kinds of pragmatic effects, the non-embeddable ones, i.e. the genuine CIs. Contrary to what he claims, a *generating* inference can be global even in cases where the pragmatic effect embeds. One of the types of cases of alleged embedded CIs discussed above provides evidence for this: the conditional case, as illustrated in (5)/(6). Let us recall that, in many (indicative) conditionals, the antecedent is naturally interpreted in a way which goes beyond its linguistically encoded content. I will focus on example (5), here repeated:

(5) A: What's making noise up in the attic?
     B: I'm not sure. But if there's a nest up there, we're going to have a big mess to clean up.
     (If there's an **occupied** nest up there...)

Let's have a look at what the triggering and generating inferences (in simplified form) would look like in this case, starting with the former:
   i. The speaker asserts the conditional *s* with literal (i.e. unenriched) content *n* e.g. If there's a nest..., ... big mess
   ii. Given that *s* is a conditional, the speaker's asserting *s* would violate the Co-operative Principle, in particular the maxim of Quality, unless she also believes some enriched version of it (as the unenriched antecedent would produce a blatantly false conditional: empty nests would cause no mess).
   iii. There is no reason to think that the speaker is not observing the Cooperative Principle.
   iv. Therefore, the conditional proposition the speaker wishes to assert is not the one corresponding to *n*.

The generating inference, as would be expected, picks up from the triggering one:
   v. Given that the conditional proposition the speaker wishes to assert is not the one corresponding to *n*, it must be a proposition which results from a minimal enrichment of *n* that is contextually adequate, in such a way that the conditional becomes true and Cooperativity is observed.
   vi. Such an enrichment produces the following strengthening of the antecedent:
   vii. n' = If there's an **occupied** nest, ...
   viii. The speaker knows that the hearer is able to go through the reasoning in (i)-(vi) and has done nothing to prevent the hearer from concluding that the speaker believes the enriched conditional proposition.
   ix. Therefore, the speaker intends the hearer to think she believes the enriched conditional proposition.
   x. Therefore, the speaker has asserted that proposition.

Again, as in the disjunctive case, in (5) the triggering inference is global: it makes crucial use of the premise that an assertion has been made – an assertion, as we have seen, whose compliance with the Cooperative Principle has to be worked out by enriching the linguistically encoded content being put forward. But it cannot be denied that the generating inference leading to that enrichment is, in this case, also global. In particular, it is apparent that step (v) crucially states the fundamental role of s's status *as a conditional* in the whole derivation so as to motivate

the derivation of a specific enrichment of the antecedent: it must be one that, besides being contextually relevant (as in the disjunction case), makes the whole conditional true (hence the generation of *n'*). In other words, both in the triggering component and in the generating component of the overall inference one needs to take into account the fact that the assertion of a conditional has taken place. In the former case one needs to realize that the utterance would not have been cooperative had the speaker not intended the antecedent to have an enriched interpretation; in the latter, one needs to realize which specific enrichment would be most cooperative. Consideration of the whole utterance is required for both inferences to go through, even if the resulting pragmatic effect does not arise (as mentioned above) from a previous act of assertion and is therefore (under the assumption that Recanati is right about NECI), not a CI.

It follows that Recanati's claim that the CI is the only pragmatic kind of inference that is fully global is false. Although full-fledged globality can reasonably be said to be a necessary condition for the generation of a CI, it surely is not a sufficient condition, contrary to what he suggests. In other words, although both the triggering and generating inferences must be computed globally for the inference to count as a CI (since there is arguably no CI without the actual assertion of the input content, and actual assertion entails full-fledged globality) it can be the case that both kinds of sub-inference are global without the resulting overall inference counting as a CI (since it can be the case that neither of them, and in particular the generating one, stems from the actual assertion of the unenriched propositional content). Thus globality does not offer a reliable criterion for distinguishing CIs from other Gricean (*latu sensu*) inferences. Furthermore, assuming (with Recanati) that a pragmatic inference is a CI just in case it is non-embeddable, the globality criterion cannot account for embeddability either. The conditional case shows that embeddable Gricean inferences (i.e. non-CI ones) can be fully global, just like CIs.

One alternative strategy for drawing the line between CIs proper and other kinds of pragmatic inference is Recanati's own well-known, although not widely popular, "Availability Criterion" or AC (cf. Recanati 2004, 2010). According to the AC, an utterance's propositional content counts as literally asserted content (and, correspondingly, as the utterance's truth-evaluable content) only if it is "available", i.e. only if it can be worked out by the parties in a conversational interchange as a recognizable element of the overall message conveyed by that utterance. This is usually taken to mean that, in a CI, the *implicans* (the literally asserted proposition) and the *implicatum* (the implicated proposition) can intuitively be distinguishable from each other and, *a fortiori*, that the former must be intuitively recognizable as the input to the latter and, thus (as required by the notion of a CI

calculus) as actually asserted. Given that embedded pragmatic effects do not comply with these requirements (in particular, given that the linguistically encoded, pragmatically unenriched, content of the utterance is arguably not intuitively recognizable as being asserted), they cannot be genuine cases of CIs, even if they can be broken down into a Gricean-style "rational reconstruction" (as Simons suggests).

The AC is, fairly obviously, a way of bringing to the fore the role of speakers' intuitions as the grounds for the theorist's verdict when deciding on what, within the realm of pragmatic inferences, deserves to be assigned the status of CI. This is, to be sure, one cause for its considerable unpopularity. Intuitions have not always enjoyed a great deal of prestige as the providers of solid foundations for theory in the philosophy of language, and the recent upsurge of the debate on their role in general philosophy has not provided much solace in that respect. Recanati's globality test would, on the face of it, be a reasonable, and more solid, substitute, in that it equates the notion of a CI with that of an embeddable pragmatic inference and verifies whether a pragmatic inference is fully global or just partially so, thus apparently setting CIs apart from other kinds of pragmatic inferences in a theoretically sounder way than provided by the AC. In particular, it avoids the somewhat crude reliance on untutored speakers' intuitions required by the adoption of the AC. Unfortunately, as we have seen, that globality test is less than reliable. Furthermore, as is also apparent from the above discussion, applying the globality criterion commits the theorist to adopting the very same AC it was supposed to be replacing. In fact, as we have seen, establishing whether a pragmatic inference is fully global, as proposed by Recanati, entails establishing whether an utterance's linguistically encoded content is actually being asserted; and the ultimate deciders of *that*, in turn, are speakers' intuitions about what counts as the utterance's literal truth-evaluable content, as recommended by the AC. So even if it worked, the globality test could not be an alternative to the AC, since it cannot be applied without adopting that very same AC.

It seems, therefore, that the AC is the ultimate warrant of a Recanatian kind of rejection of the CI status to embeddable pragmatic inferences. The metaphilosophical question of ascertaining how awkward this result is goes beyond the scope of this paper.

References

Anscombre, Jean-Claude/Ducrot, Oswald (1983): *L'Argumentation dans la Langue*. Brussels: Mardaga.

Bach, Kent (1994): "Conversational Impliciture". In: *Mind and Language* 9(2), pp. 124–162.

Borg, Emma (2012): *Pursuing Meaning*. Oxford: Oxford University Press.

Chierchia, Gennaro (2004): "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface". In: Adriana Belletti (Ed.): *Structures and Beyond*, New York: Oxford University Press, pp. 39–103.

Cohen, L. Jonathan (1971): "The Logical Particles of Natural Language". In: Yehoshua Bar-Hillel (Ed.): *Pragmatics of Natural Language*, Dordrecht: Reidel, pp. 50–68.

Green, Mitchell (1998): "Direct Reference and Implicature". In: *Philosophical Studies* 91, pp. 61–90.

Geurts, Bart (2009): "Scalar Implicatures and Local Pragmatics". In: *Mind and Language* 24, pp. 51–79.

Levinson, Stephen (2000): *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.

Recanati, François (2003): "Embedded Implicatures". In: *Philosophical Perspectives* 17, pp. 299–332.

Recanati, François (2004): *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, François (2010): *Truth-Conditional Pragmatics*. Oxford: Clarendon Press.

Recanati, François (2017): "Local Pragmatics: Reply to Mandy Simons". In: *Inquiry* 60(5), pp. 493–508.

Simons, Mandy (2010): "A Gricean View on Intrusive Implicatures". In: Klaus Petrus (Ed.): *Meaning and Analysis: New Essays on Grice*. Palgrave-Macmillan, pp. 138–169.

Simons, Mandy (2014): "Local Pragmatics and Structured Contents". In: *Philosophical Studies* 168 (1): 21–33.

Simons, Mandy (2017): "Local Pragmatics in a Gricean Framework". In: *Inquiry* 60(5), pp. 466–492.

Palle Leth

Is There any Use for a Notion of the Correct Interpretation of an Utterance?

Abstract: Philosophers of language and theoretical linguists disagree about whether the meaning of an utterance is determined by the speaker's more or less constrained communicative intention or by public features available to the hearer such as conventional linguistic meaning and diverse contextual cues. In this paper I question the common presupposition that there is such a thing as the correct interpretation of an utterance. I set out to show that, from the viewpoint of the three options at stake in the interpretive interaction between the speaker and the hearer, the notion of correct interpretation is dispensable. If the hearer takes an interest in the speaker's intended meaning, as she most frequently does, she has no reason to be concerned with the actual meaning of the utterance. If the hearer wants to hold the speaker responsible for the normative consequences of the utterance, what is at issue is whether the hearer had the best reasons to take the utterance the way she did, i.e. the hearer opts for the most reasonable interpretation of the utterance, irrespective of the speaker's intention and also of the objective meaning of the utterance. Finally, the hearer may opt for a merely imagined meaning of the utterance, which accounts for the remaining purposes a hearer can have with respect to an utterance, e.g. merriment. If there is no use for the notion of the correct interpretation of an utterance in interpretive practice, it seems that there are good reasons to dispense with this notion also in the theory about language and communication.

Keywords: utterance meaning, utterance interpretation, what is said, intentionalism, anti-intentionalism, contextualism, semantics/pragmatics, accountability

1 Introduction

Many philosophers of language and theoretical linguists take it as a matter of course that there is such a thing as the correct interpretation of an utterance. When a speaker utters a certain sentence with a certain communicative intention and in a certain context, it is thought that there is something that determines what the meaning of the utterance is or what the speaker counts as having said.

Most theorists are here concerned with the semantic, propositional, or truth conditional component of utterance meaning, which is thought to constitute what is actually or objectively said by the utterance at a given occasion. From the perspective which I will adopt here, however, there is no reason not to treat the determination of pragmatic components of utterance meaning on a par with the determination of semantic components.

There are basically two approaches to determining utterance meaning. For intentionalists it is the speaker's intention, subject to more or less severe constraints, which constitutes the meaning of an utterance. For anti-intentionalists it is the public features available to the hearer which settle the meaning of an utterance. Most theorists thus take for granted that there are three entities at stake: the speaker's intention, the hearer's interpretation, and the meaning of the utterance. The question is basically whether utterance meaning is to be modelled upon the speaker's intention or the hearer's interpretation. The debate has generated a large number of sophisticated accounts, several of which combine ingredients from both of the main approaches. But as the conception of the criteria of utterance meaning to a large extent relies partly on intuitions concerning invented scenarios and partly on quite general assumptions about the roles of the speaker and the hearer in linguistic communication, it is not surprising that there is as yet no consensus as to what constitutes the correct interpretation of an utterance.[1]

I will not enter into the details of this debate here. Instead, I will question the common presupposition that there is such a thing as the correct interpretation of an utterance.[2] If we consider the interpretive interaction between the speaker and the hearer, there seem to be three approaches which a hearer can take with respect to a speaker's utterance: the hearer may take an interest in the speaker's intended meaning or in the most reasonable meaning of the utterance or in a

---

1 Abstracting from important differences in scope and emphasis, for some anti-intentionalist accounts (see, e.g., Kaplan 1978, McGinn 1981, Wettstein 1984, Dummett 1986, Travis 1989, Reimer 1991, Green 2001, Saul 2002, Gauker 2008, Mount 2008; for some intentionalist accounts, see, e.g., Davidson 1986, Kaplan 1989, Neale 1992, Predelli 1998, Recanati 2005, Perry 2009, Korta & Perry 2010, Montminy 2010, Stokke 2010, Åkerman 2010, King 2013, Neale 2013, King 2014, Speaks 2016, Neale 2016, Speaks 2017).
2 For theorists to a smaller or larger extent sceptical about objective determination of (aspects of) utterance meaning (see, e.g., Ariel 2002a, Ariel 2002b, Pagin 2002, Smit 2012, Heck 2014, Gauker 2015, Åkerman 2015, Bach 2017). Most of these theorists are concerned with demonstrative reference only. I differ from them in attempting a more comprehensive account, dispensing with objective utterance meaning across the board, taking into consideration what I believe is the complete variety of interpretive options and in grounding the account in the interpretive interaction between the speaker and the hearer.

merely imagined meaning of the utterance. I will argue that the first option excludes taking any interest in utterance meaning (section 2); the second option is sufficient for accountability purposes (section 3); and the third option suffices for the remaining purposes a hearer can have with respect to an utterance (section 4). If there is no practical utility in the notion of the correct interpretation of an utterance, this seems to speak strongly in favour of its being dispensed with also in theory.

2 Alice and intentions

One major reason for being an intentionalist about utterance meaning is the natural assumption that language is in the service of communicating beliefs and desires. The goal of interpretation ought thus to be the hearer's discovery of the speaker's communicative intention. Taking for granted that there is such a thing as utterance meaning, one is hereby led to say that the speaker's intention is what settles utterance meaning. It is certainly true that the hearer very often takes an interest in what the speaker means; this seems indeed to be the hearer's default approach to an utterance. The hearer's interest in the speaker's intention is, for example, manifested by her asking for the speaker's meaning in case she does not get at it.

Taking an interest in the speaker's intention appears, however, to be as simple as making a distinction between what the speaker wants to say and what her utterance means. For the meaning of an utterance cannot be just anything the speaker happens to intend to convey or pretends to intend to convey by her utterance. The intentionalist conception of utterance meaning is indeed challenged by Humpty Dumpty, the nursery rhyme character who in Carroll's second Alice book claims that his utterance means just whatever he chooses it to mean. From the viewpoint of an intentionalist conception of utterance meaning, according to which an utterance means what the speaker means, it is necessary to point out the limits of what the speaker can mean by her utterance.

In this section I will argue that the hearer's interest in the speaker's intention, far from lending support to the view that the speaker's intention settles the meaning of an utterance, actually excludes taking an interest in positing or establishing utterance meaning. Once the presupposition that any such thing as utterance meaning should be settled, Humpty Dumpty ceases to be a problem. Humpty Dumpty is not vindicated, but disarmed, and the importance of the speaker's intended meaning is in no way diminished for our giving up (its determining) utterance meaning.

2.1 What a speaker can mean

In her well known exchange with Humpty Dumpty, Alice situates his meaning pretentions in the following manner:

> The question is [...] whether you *can* make words mean so many different things.
>
> (Carroll 1872, p. 251)

Humpty Dumpty had concluded a piece of reasoning on the number of unbirthday presents you can get with an utterance of the sentence *There's glory for you!*, by which he – to take his word for it – means 'There's a nice knock-down argument for you!'.

Alice's implied verdict is understandable. As far as linguistic conventions are concerned, the word *glory* does not mean 'a nice knock-down argument'. And in so far as Humpty Dumpty does not provide any alternative clues as to his alleged meaning, it seems reasonable to hold that he has not managed to say what he pretends to have attempted to say. Alice's verdict is shared by many theorists. Here is, e.g., what Davidson says:

> Humpty Dumpty is out of it. He cannot mean what he says he means because he knows that "There's glory for you" cannot be interpreted by Alice as meaning "There's a nice knock-down argument for you". We know he knows this because Alice says "I don't know what you mean by 'glory'", and Humpty Dumpty retorts, "Of course you don't – till I tell you".
>
> (Davidson 1986, p. 98. Cf. MacKay 1968, p. 200)

There are perhaps things which Humpty Dumpty can mean by his utterance of "There's glory for you!" and other things which he cannot mean by it, and perhaps he is or pretends to be confused about that. But actually it seems to me that the real question to be asked is for what reasons the question of what Humpty Dumpty can mean by his utterance or what the meaning of the utterance is becomes an issue at all.

Alice's and Humpty Dumpty's exchange presumably instantiates the following pattern. The hearer is confronted with an utterance of the speaker, which the hearer cannot understand. The hearer's non-understanding is related to what the speaker means in the sense of what the speaker is getting at. It is possible that the hearer understands the sentence at its purely linguistic level perfectly well; it is also possible that she is able to recover the material which is necessary to resolve all sorts of semantic underdetermination, like ambiguity, reference, etc. But as long as the processes of decoding and inferencing do not permit the hearer to make sense of the utterance, she will be dissatisfied and in an intuitive and im-

portant sense she will not understand the utterance. In case the author of the utterance is near at hand, it is very natural for the hearer to ask the speaker what she meant. In very many cases the speaker will provide the hearer with a novel utterance which the hearer will understand to her own and the speaker's satisfaction, or the speaker will provide the hearer with an additional utterance which together with material from the original utterance permits the hearer to grasp what the speaker wanted or wants to convey.

It is at this juncture that some hearers and theorists feel the urge to compare the speaker's alleged intention and her original utterance and occasionally come up with the verdict that the speaker could not mean or could not have meant what she claims to mean or have meant. Such a verdict could be based on the fact that, according to the rules of the language, there is no way from the linguistic meaning of the sentence uttered to the allegedly intended meaning. Or it could be based on the fact that the speaker did not provide any contextual clues as to her intended meaning. Some theorists claim that there is a connection between linguistic facts about lexicon and syntax and utterance meaning so that what is said is constrained by linguistic or compositional meaning. Others think that there is a metaphysical connection between what the hearer can understand and what the speaker can mean in such a way that the speaker could not have meant something which the hearer could not have understood.[3] But even if there were such connections, I cannot see how they could pertain directly to the theory of utterance interpretation.

2.2 Taking Humpty Dumpty's point

In order to make this claim, I will, as it were, invoke Humpty Dumpty's authority.[4] That there is no reason for Alice to pronounce on what he can mean by his utterance seems precisely to be Humpty Dumpty's implicit point. Humpty Dumpty initially says to Alice:

---

3 For proponents of blocking Humpty Dumpty by conformity to conventional linguistic meaning, see, e.g., Dummett (1986), Kotátko (1998) and Green (2001); for proponents of blocking Humpty Dumpty by constraints on intention formation, see, e.g., Donnellan (1968), Davidson (1986) and Neale (2016).
4 Or Dodgson's: "[A]ny writer of a book is fully authorised in attaching any meaning he likes to any word or phrase he intends to use. [...] I meekly accept his ruling, however injudicious I may think it" (Dodgson 1896, p. 166).

> There's glory for you!

> (Carroll 1872, p. 251; as the following quotes)

Alice's immediate reply is:

> I don't know what you mean by 'glory'[.]

Alice thereby indirectly asks Humpty Dumpty for his intention, as if it were a matter of his giving her information about it. She seems indeed to give Humpty Dumpty *carte blanche* as to his meaning. In general, when one asks for information, she is not already in possession of it herself. It is natural not to know what one asks to know. It is therefore not altogether inept on Humpty Dumpty's part to respond:

> Of course you don't – till I tell you.

Humpty Dumpty then promptly delivers the required information:

> I meant 'there's a nice knock-down argument for you!'

It is at this point that Alice remonstrates him. Her immediate reaction to his clarification is to object:

> But 'glory' doesn't mean 'a nice knock-down argument'[.]

Certainly it does not, but since when is the issue what *glory* means? Alice's initial question seems to be motivated by her not understanding Humpty Dumpty's intended meaning and by her wanting to know what he wants to convey by his utterance. If Alice is interested in the lexical meaning of *glory*, as she now seems to be, there is no reason to ask Humpty Dumpty about *glory* at all, since Alice herself seems to be perfectly knowledgeable about its meaning. Perhaps there is a tacit expectation on Alice's part that Humpty Dumpty's meaning of *glory* should fall within the linguistically licensed sphere of meaning of the word *glory*. Some theorists think that the speaker's intention determines the meaning of her utterance as long as it is in conformity with conventional linguistic meaning. Such a position makes apparent sense in theory, but a request in the form of 'Tell me what you wanted to say, provided that it be something expectable!' makes definitely less sense. Such a restricted interest could hardly qualify as a genuine interest in the speaker's intention. In the face of Alice's somewhat feigned interest in his intention, Humpty Dumpty's recourse to his strained theory of word meaning is perhaps understandable:

> When *I* use a word, it means just what I choose it to mean – neither more nor less.

It is certainly not the case that a word means what the individual speaker wants to convey by it on a particular occasion. But, on a particular occasion, it seems that the question is not about what words mean, but about what speakers mean. That seems to be what Humpty Dumpty somewhat hyperbolically is suggesting. Alice rebukes:

> The question is [...] whether you *can* make words mean so many different things.

This is precisely not the question, on a charitable construal of Humpty Dumpty's arrogance. For Alice seems to be motivated by two mutually exclusive interests: an interest in Humpty Dumpty's intention and an interest in what he can mean by his utterance. If she takes an interest in Humpty Dumpty's intention, there is no reason to be interested in the possible meaning of the utterance, and if she takes an interest in the possible meaning of the utterance there ought to be no interest in Humpty Dumpty's intention. Any interest in the word meaning of *glory* should be eclipsed by Alice's avowed interest in Humpty Dumpty's intended meaning. No particular purpose is served by Alice's verdict that Humpty Dumpty could not have meant what he pretends to have meant as long as she is in any case asking for and prepared to go along with his intended meaning.

This does not, of course, prevent Humpty Dumpty from being rightfully accused of not speaking intelligibly. But that is not *eo ipso* a reason to pronounce on what he can mean by his utterance. That might, however, be a good reason to stop talking to him. That is why Humpty Dumpty answers:

> The question is which is to be master – that's all.

For the speaker it is all about making the hearer interested in her intention, i.e. to be master. Humpty Dumpty is master for yet a little while. He continues:

> They've a temper, some of them – particularly verbs: they're the proudest – adjectives you can do anything with, but not verbs – however, *I* can manage the whole lot of them! Impenetrability! That's what *I* say!

And Alice asks, in a more submissive tone:

> Would you tell me, please, what that means?

Humpty Dumpty answers:

> Now you talk like a reasonable child[.]

He then sets out to explain – I would not say what that meant nor what he meant, but – what he wanted to convey or pretends to have wanted to convey.

Is not the difference between an actual intention and a simply alleged intention crucial? If Humpty Dumpty is not serious himself, there is certainly no reason for us to take him seriously. I am not certain that it is easily settled that Humpty Dumpty is not serious and furthermore I do not think that it makes an important difference whether he is or not. First, I am not sure that 'There's a nice knock-down argument for you!' does not constitute a reasonable interpretation of 'There's glory for you!', there being cues permitting Alice and other hearers to get at this interpretation. Second, when it comes to speakers who preface their every utterance with a select piece of deliberate gibberish, I think the main objection is not that they have illegitimate meaning pretentions, but simply that they are tiresome. The main issue for us is not whether they can mean what they pretend to mean, but whether we should pay attention to them or not. That we have good reasons not to understand the gibberish is a different matter, which I will take up in section 3.

The concern with what speakers can mean by their utterances has its roots in an intentionalist conception of utterance meaning. It is from the viewpoint of such a conception that Humpty Dumpty constitutes a challenge and needs to be blocked: the meaning of an utterance cannot be just anything the speaker intends. But the two ways of blocking Humpty Dumpty, conformity to linguistic conventions or constraints on intention formation, do not have any direct implications for utterance meaning. Considerations about linguistic meaning or intention formation are important factors in the hearer's hypothesizing about the speaker's intention. Given the fact that *glory* doesn't mean 'knock down argument' or the possible fact that Humpty Dumpty does not provide any contextual cues in the direction of his meaning 'knock down argument' by *glory*, it is unlikely that he wants to convey 'knock down argument' by his use of *glory*. But without any further stipulation about the relation between such facts and utterance meaning, those facts do not show anything beyond that. Such facts by themselves show at most that the speaker spoke in a way she should not have spoken for the purpose of being understood or that the hearer had very good reasons not to understand the utterance as intended.

2.3 Opting for intended meaning

One fundamental attitude in utterance interpretation is taking an interest in the speaker's intended meaning. This interest excludes interest in such a thing as the correct interpretation of the utterance. As long as the hearer is prepared to go

along with the speaker's intended meaning, there is no reason to posit or establish utterance meaning. This is what we will now see in various less-extreme cases than Humpty Dumpty.

Let us first consider cases of communicative success, i.e. cases in which the hearer understands the speaker's utterance to the speaker's satisfaction. Many theorists account for communicative success by positing an entity which both the speaker's intention and the hearer's interpretation coincide with, namely the meaning of the utterance. But it is quite clear that successful communication does not require more than the simple convergence of the speaker's intention and the hearer's interpretation.[5]

In cases where the linguistic and contextual material presented to the hearer is underdetermined and the speaker's intention is compatible with it, many theorists say that the speaker's intention determines the meaning of the utterance. Meaning is up to the speaker in cases of ambiguity, reference, assertions, and implicatures. As MacFarlane says with reference to Nunberg's ham sandwich metonymy (Nunberg 1995):

> It seems to me that if we wanted to settle, for example, whether Nunberg's waitress had asserted that a *sandwich* had left, or that a *person* who ordered a sandwich had left, we might ask which (if either) of these propositions she meant to commit herself to. To answer this question is to settle what she asserted.
>
> (MacFarlane 2011, p. 92)

I think that a convention-abiding speaker's intention settles utterance meaning just as little as Humpty Dumpty's wayward intention. For how could the speaker's answer to our question of what she meant to commit herself to be said to settle what she asserted or, more generally, the meaning of the utterance? In order to make good such a claim, I believe that the theorist has to show at least three things.

First, it must be shown that the hearer, in asking for the speaker's intended meaning, is really asking for the meaning of the original utterance and not for a novel utterance which she will more easily understand. I think that an elucidation of meaning is rather to be construed as an erasure of the original utterance and its replacement by the novel utterance. It is true that many clarifications take an elliptical form and must be understood together with the original utterance,

---

5 This point is explicitly made in Neale (2007). However, Neale thinks nevertheless that it is the speaker's intention which determines utterance meaning at the metaphysical level. I see no reason for singling out the speaker's intention in this way nor for not dispensing with utterance meaning altogether.

but that does not prevent us from taking the meaning of the novel utterance to be what is fundamentally at stake. Second, it must be firmly established that the speaker in reporting her original intention is reliable, for a simply alleged intention cannot be said to constitute the meaning of an utterance. The speaker may be oblivious, confused or insincere. In most cases, I presume, the hearer does not care to ascertain that this is the case and as long as she does not, I cannot see how one can support the idea that asking for intentions settles utterance meaning. The fact that the hearer does not care about ascertaining that the speaker is reliable suggests that, rather than asking for factual information about the meaning of the original utterance, she is inviting the speaker to make a novel utterance. Third, even if the hearer is firmly convinced that she is concerned with the meaning of the original utterance and has ascertained that the speaker is reliable, she must somehow make sure that she takes the speaker's intention to be decisive for some strictly principled reasons and not merely because she is invited by the speaker to go along with the intended meaning. How could the hearer make clear that she is concerned with the meaning of the original utterance and not with the speaker's intended meaning? For these reasons, I believe that the view that the original utterance is, as it were, erased and replaced by a novel utterance is simple and sufficient.

I think the same account holds for the cases of confusion which are so frequently discussed in the literature. Here is one of the scenarios:

> Suppose that without turning and looking I point to the place on my wall which has long been occupied by a picture of Rudolf Carnap and I say:
> (1) [That] is a picture of one of the greatest philosophers of the twentieth century.
> But unbeknownst to me, someone has replaced my picture of Carnap with one of Spiro Agnew.
>
> (Kaplan 1978, p. 30; layout added)

Intentionalists worry about cases like this, because it may seem as if the intentionalist account is contradicted by such cases. Surely, the speaker has not referred to the picture of Carnap in virtue of her intention to refer to it, if she, as a matter of fact, points at a picture of Agnew? In order to circumvent any anti-intentionalist conclusion, intentionalists point out that speakers typically have a set of intentions when making utterances. In this case, the speaker has conflicting intentions. The speaker has the intention to refer, on the one hand, to the picture of Carnap which is on her mind and, on the other hand, to the picture which is on the wall behind her (cf. e.g. Bach 1987, Perry 2009 and King 2013). It is the latter intention – proximal or directing – which is claimed to determine the referent of the demonstrative in the utterance.

If we look at this scenario from the speaker's and the hearer's point of view, I think that the problem evaporates. The hearer is confronted with evidence which goes in diverging directions. It seems most plausible that the hearer will handle this confusion in the same way that she handles ambiguity and other cases of underdetermination: she will ask the speaker what she wanted to refer to and go along with the answer. And even in cases where she, for some reason or other, is not in a position to ask the speaker, the question which she will ask herself is what the speaker wanted to say and not what the meaning of the utterance was. The conflicting intentions account seems to me to provide an answer to a question which there is no particular reason to ask.

2.4 Conclusion

I have so far argued that the hearer's taking an interest in the speaker's intention excludes taking an interest in positing or establishing any such thing as the meaning of an utterance. There are facts about which meaning is licensed by the rules of language and facts about the availability of adequate cues from the hearer's point of view, but those facts do not have any implication for such a thing as utterance meaning except by stipulation. No such stipulation is called for. There is no reason why compositionality or linguistic competence should take us further than sentence meaning at most. From the viewpoint of the hearer's interest in the speaker's intention, there is no reason to tell whether the speaker's intention corresponds or does not correspond to any such thing as the meaning of the utterance. Humpty Dumpty, for all his arrogance, is basically in the right: if the question is what the speaker's intention is, there is no place for the notion of correct interpretation. For either the hearer finds out what the intended meaning is and then there is no further question, or the hearer does not find it out and then the question is what it is, not what the correct interpretation is.

3 The Mayor of Clay and accountability

Taking an interest in the speaker's intention is not the hearer's only option in utterance interpretation. The speaker is not always the master. One major reason for being an anti-intentionalist about utterance meaning is the phenomenon of accountability. A speaker may be accountable for her utterance, i.e. responsible for the normative consequences of her utterance. Most theorists think that ac-

countability does not depend on the speaker's subjective intention, but on objective features publicly available to the hearer, i.e. linguistic conventional meaning and contextual cues such as demonstrations, salience, discourse topic, common ground. Taking the existence of utterance meaning for granted, many theorists thus take utterance meaning to be settled on anti-intentionalist grounds. But the invocation of utterance meaning for the purposes of utterance responsibility seems unnecessary. When accountability is at stake, the question we ask or should ask is solely whether the hearer had the best reasons to take the utterance the way she did, not what the meaning of the utterance was. If epistemic evaluation of the speaker's interpretation is sufficient for matters of responsibility, another motive for positing or establishing utterance meaning vanishes.

3.1 Taking the speaker to task

In November 2016, more than 175,000 people petitioned for the resignation of the Mayor of Clay County in West Virginia, U.S.A. The reason was that the Mayor had uttered the sentence *Just made my day, Pam*. The utterance was made in connection with a friend's Facebook statement arguably containing a racist comparison. Part of the text of the petition reads as follows:

> Clay County Development Corp. Director Pamela Taylor Ramsey posted a hateful, racially motivated statement to Facebook, which was not only liked by the Clay County Mayor Beverly Whaling, but commented on in a way that showed she agreed with her comments.
> URL: http://www.thepetitionsite.com/sv/takeaction/571/113/900/.

We can see that the author of the petition thinks that she is entitled to pronouncing on the Mayor's state of mind, the Mayor's being said to have agreed with the comments of her friend. Among the comments made by the petitioners, one may furthermore find claims to the effect that the Mayor's utterance of 'Just made my day, Pam' was racist, as in this statement:

> Making racist remarks is not how we do things in [C]lay [...].
> URL: http://www.thepetitionsite.com/sv/takeaction/571/113/900/

And also claims to the effect that the Mayor is a racist person on account of this utterance:

> She is clearly a racist!!!!
> URL: http://www.thepetitionsite.com/sv/takeaction/571/113/900/.

In a statement to the Washington Post, the Mayor sponsored an answer to the question of what she meant to commit herself to:

> My comment was not intended to be racist at all. I was referring to my day being made for change in the White House! I am truly sorry for any hard feeling this may have caused! Those who know me know that I'm not of any way racist! Again, I would like to apologize for this getting out of hand!
> URL: https://www.washingtonpost.com/news/post-nation/wp/2016/11/14/ape-in-heels-w-va-officials-under-fire-after-comments-about-michelle-obama/?utm_term=.f24d17261f79

We can see that the mayor denies, in turn, that she had a racist intention when making the utterance; that her utterance was racist – because, as a matter of fact, it referred to a non-racist part of the friend's post; that she herself is a racist person.

On the one hand, the petitioners' uptake seems to be concerned with both the speaker's intended meaning and with the meaning of the utterance. On the other hand, we have the Mayor's denial of any racist speaker intention or utterance meaning. Disagreement on these issues did not prevent the Mayor from resigning shortly after the petition was published. This seems to indicate that, despite appearances, neither the Mayor's intention nor the meaning of her utterance were involved in the petitioners' taking the Mayor to task for her utterance and in her resigning.

3.2 Opting for most reasonable meaning

The sentence *Just made my day, Pam*, as far as its linguistic meaning goes, does not contain any racism. The utterance of that sentence can only be taken as racist if it is read as a comment of the friend's Facebook post and furthermore as agreeing with a specific part of it. The alleged racism of the utterance of 'Just made my day, Pam' is conveyed by other means than purely linguistic; it is implicit and has to be inferred. According to many philosophers, in cases of linguistic underdetermination such as disambiguation, reference, assertion, implicatures, innuendos, the hearer submits to the speaker's authority. Philosophers typically claim that what characterizes such aspects of utterance meaning is their deniability. Witness, e.g., Camp commenting on Grice's recommendation letter case (Grice 1967):

> By expressing herself so indirectly, the speaker avoids going on record as actually *saying* anything mean about Mr. X: she cannot be quoted or otherwise cited as saying anything unkind. Someone might object that given the context of utterance, it was perfectly obvious that the speaker meant that Mr. X is a bad philosopher. However, while this might well be

obvious for many practical purposes, it is still open to the speaker to respond that it only appeared obvious to the hearer because he made additional assumptions about her communicative intentions, and to insist that those assumptions were unwarranted in this case. Indeed, it is precisely the desire to preserve the option of such a response that presumably led the speaker to express herself in such a roundabout way in the first place. While such a response may be disingenuous, it is one that politicians and diplomats offer regularly, and one that is not available in cases where the speaker has directly and literally stated her intended meaning [...]. Thus, implicatures offer a brand of communication with deniability.

(Camp 2008, p. 8. Cf. also Bach & Harnish 1979.
For an exception, see Saul 2002)

If this were true, the petitioners could only set out to prove the Mayor wrong about her communicative intention. To judge from the comments this is also something some of them attempt to do. But actually, they do not have to engage in this difficult task. For, contrary to what most philosophers of language assume, the realm of the implicit is not up to a speaker's denial, and the speaker's accountability goes far beyond literal meaning. Courtroom practice, and also ordinary practice, differ significantly from philosophical orthodoxy. In cases of offensive utterances, courts generally take a speaker to be responsible for what-ever meaning the hearer had the best reasons to take the speaker to convey. The innocence of the literal meaning of the sentence does not always disculpate the speaker from the offensiveness caused by the implicature, which hearers had good reasons to take the speaker to make by her utterance. Here is an example of what legal scholars say:

In deciding what words mean for the purpose of defamation, the intention of the writer or speaker is largely irrelevant. The test is the effect on the ordinary reader, who is endowed for this purpose with considerable wisdom and knowledge of the way of the world. The literal meaning is not conclusive: the ordinary reader knows all about irony.

(Robertson & Nicol 2002, p. 82. Cf. Durant 1996
and Shuy 2009)

And here is another statement to the same effect:

A statement does need not make a direct criticism in order to be defamatory – a defamatory implication or innuendo can be just as dangerous. [...] the courts will look at what the ordinary, reasonable reader would think that the words implied. [...] In general, in defamation the important issue is not what you meant to say, but what a jury believes a reader would

think you meant. So it is no defence to say that you did not mean to defame the claimant, or did not realise what your words implied.

(Quinn 2015, pp. 214, 214 and 216, respectively.
Cf. Smartt 2006)[6]

The evidence from court practice makes clear that the speaker's intention is not always the master. Whether a speaker is convicted of defamation, for instance, depends not on the speaker's intention, but on the reasonableness of the hearer's interpretation. Legal reasoning is anti-intentionalist in this respect. It certainly happens that judges take themselves to be establishing the meaning of an utterance as if their verdict parted from the correct interpretation of the utterance. But this assumption is actually inessential. To take the most reasonable interpretation of an utterance to constitute the correct interpretation of the utter-ance involves judges and hearers and anti-intentionalists in unnecessary dispute with defendants and speakers and intentionalists.

Tying liability to the most reasonable interpretation of an utterance is not unique to judges. Remember that the Mayor of Clay herself apologized 'for this getting out of hand'. So she herself assumed responsibility for the interpretation made by the hearers, though it did not at all, according to her, correspond to her intended meaning. The petitioners need only claim that the consequences of the Mayor's responsibility should go further than mere apology.

In matters of accountability, any claim as to the actual meaning of an utterance is better reconstructed as a claim as to the reasonableness of the hearer's interpretation. When a hearer uses a locution such as 'That's certainly what you wanted to say, but that's not what you actually did say', she should be understood as doing no more than just emphasizing her good reasons to take the utterance the way she did. The invocation of the notion of utterance meaning is wholly dispensable. A simple notion of the most reasonable interpretation suffices for matters of accountability.

---

6 This is not to say that whenever a defamatory interpretation is possible it will be adopted by the jury. Some courts practice charity: '[s]ome states, such as Ohio, have an "innocent construction rule," requiring the courts to give allegedly defamatory statements an innocent meaning whenever they are reasonably capable of an innocent meaning. If a statement is susceptible to two meanings, one allegedly defamatory and one benign, the defamatory meaning should be rejected and the benign meaning adopted.' (Shuy 2009, p. 116).

3.3 Are good reasons sufficient?

For some people and theorists I suspect that this conception of the speaker's responsibility is not strong enough. They are not satisfied with the hearer's having the best reasons to take the utterance the way she did, as if accountability were nothing but a matter of subjective point of view; they want the hearer's interpretation to have a more objective status. What animates many anti-intentionalist accounts of utterance meaning is the thought that the speaker's

> utterance [that *p*] commits her to believing that *p* in an *objective* sense, irreducible to a feeling of obligation on the part of the speaker and corresponding expectations on the part of the audience[.]
>
> (Kotátko 1998, p. 236)

Imagine for instance that the speaker and the hearer are under the illusion that a given slur is no more offensive than its neutral counterpart is for ordinary speakers and hearers. We overhear their conversation. Is this not a case where the speaker has, in an objective sense, said something offensive, irrespective of the happy communication between her and her hearer? A theory of utterance interpretation where the notion of utterance meaning or the correct interpretation of the utterance is dispensed with might seem incapable of accounting for the offense brought about or of taking it seriously enough.

But neither the speaker's allegedly innocent intention nor the original hearer's allegedly faulty interpretation prevents other hearers from taking the speaker to task. It suffices that other hearers have the best reasons to take the utterance as offensive. Whether they have is a matter of discussion of course. Should it be taken into account that the utterance was made to a single hearer who happened to share the speaker's idiolect and not to the general public which happened to overhear it? Such a question is precisely of the kind which affects our conception of the most reasonable interpretation and which may be debated in court.

The search for more objectivity than what reasonability represents seems to me to be misguided for two reasons. First, I cannot see how any objectivity in this regard can be had. Any stance on the issue of what was objectively said by a speaker's utterance will necessarily be an arbitrary decision. It cannot be settled whether it is the speaker's intention or the hearer's reasonable interpretation which constitutes the meaning of an utterance, for there are no theory-independent facts to be matched. Second, and more importantly, I cannot see any reason for having a notion of what was objectively said. Legal practice is certainly not

legislative for philosophical theory of language, but if the most reasonable interpretation is robust enough not to allow speakers to get away with it in court, the desire for more objectivity does not seem to serve any particular purpose.

3.4 What is most reasonable?

Since utterance meaning can be dispensed with for responsibility purposes, theorists had better stop arguing about whether it is the speaker's reasonably-formed intention or the hearer's reasonable interpretation which constitutes the correct interpretation of the utterance. But referring matters of utterance accountability to the most reasonable interpretation of the utterance will not put an end to possible disagreement of course. The question of what the most reasonable interpretation of an utterance is may appear to be just as vexed as the question of what the correct interpretation of the utterance is. I will not here enter into a discussion of what constitutes the most reasonable interpretation of an utterance. I think that it is an issue which to a large extent cannot be abstracted from the consideration of particular cases. I will just make a few remarks as to the form of the most reasonable interpretation.

I think it facilitates matters that the speaker's intention is firmly put to place. This does not imply that reasoning about the speaker's intention is unimportant. What the most reasonable interpretation is in most cases certainly depends on what it is reasonable to suppose that the speaker intended to convey. The speaker's actual intention as such, however, is irrelevant. A speaker may try to vindicate her intended meaning, but since what the most reasonable interpretation is depends on which cues were available to the hearer, the speaker will then have to show that her intended meaning was supported by linguistic meaning or that there were other cues available to the hearer which ought to have been taken into account for the interpretation of her utterance. The hearer, for her part, cannot just rely on what she, as a matter of fact, took the utterance to convey. The speaker is legitimately not willing to take responsibility for any whimsical interpretation of her utterance. The hearer must therefore ultimately be able to show that any reasonable interpreter in her situation would have taken the utterance the way she did. Linguistic meaning is here an important factor alongside diverse contextual cues. Also the hearer's subjective point of view may reasonably be taken into account, in so far as the hearer was the addressee of the utterance. So even though I do not think that any such thing as the correct interpretation of an utterance can be objectively established, I do think that what the most reasonable interpretation of an utterance is can be established with a sufficient amount of

objectivity. It is neither a factual matter nor an easy matter; it is a matter of rational negotiation, discussion, argument (cf. Clark & Wilkes-Gibbs 1990, Clark 1996, Ariel 2002a), which is not, however, without prospects of possible agreement.

3.5 Conclusion

In this section I have argued that the speaker's responsibility is tied to the most reasonable interpretation of the utterance. In order to hold the speaker responsible for her utterance, it suffices for a hearer to point at her good reasons for taking the utterance the way she does, irrespective not only of the speaker's intention, but also of any such thing as the meaning of the utterance.

4 Mrs. Malaprop and merriment

I have so far considered the hearer's opting for the speaker's intended meaning, which I have argued excludes interest in utterance meaning, and the hearer's opting for the most reasonable meaning of the utterance, which I have argued is sufficient when holding the speaker responsible for her utterance. But there are still other cases where the hearer may have the intuition that something like utterance meaning is at stake. I am here thinking of cases where the speaker mis-speaks for various reasons, e.g. slips of the tongue, malapropisms and all sorts of confusion. Such cases may be invoked in order to support the idea that there is such a thing as utterance meaning after all, irrespective of the speaker's intention and of the most reasonable interpretation of the utterance. It is my contention that these cases may be explained by our having recourse to the notion of imagined meaning only.

4.1 Malapropisms

Let us first consider malapropisms. Here is part of Mrs. Malaprop's discourse on a young woman's education:

> [B]ut above all, Sir Anthony, she should be mistress of orthodoxy, that she might not mis-spell, and mis-pronounce words so shamefully as girls usually do; and likewise that she might reprehend the true meaning of what she is saying.
>
> (Sheridan 1775, p. 86 [I.2])

Some theorists would certainly say that our merriment comes from the fact that Mrs. Malaprop actually says something which she clearly does not want to say (e.g. Reimer 2004). If so, it is the contrast between actual utterance meaning and intended meaning which explains our merriment. But do we need to invoke the notion of utterance meaning in order to account for the mis-speaking and our merriment? I believe that our merriment is perfectly explicable with recourse to linguistic and imagined meaning only.

Orthodoxy means 'orthodoxy' according to the dictionary, but the contextual cues which the hearer also considers in coming up with a hypothesis about Mrs. Malaprop's intention make it plausible that she is confusing *orthodoxy* and *orthography*. Thus the hearer gets at the point which Mrs. Malaprop wants to make, while also thinking that she has a funny way of putting it. The contrast which is involved here seems to be between Mrs. Malaprop's intended meaning and the linguistic meaning of the terms or the sentence she uses. There is no reason to take anything to be the meaning of the utterance.

As to Mrs. Malaprop's second malapropism in the quote above, there is perhaps a stronger case to be made for there being a contrast between what the speaker wants to say and what she actually does say. What is fun may seem to be that Mrs. Malaprop wants to say that a girl should apprehend the true meaning of what she is saying, but that she happens to assert that she should reprehend the true meaning of what she is saying.

But on what grounds would 'reprehend the true meaning of what she is saying' be considered to be the true meaning of Mrs. Malaprop's utterance? Given the contextual cues at hand, we are convinced that that is not Mrs. Malaprop's intended meaning nor the most reasonable interpretation of her utterance. It would be in virtue of the rules of language that this meaning should be considered to be the meaning of the utterance. But then we would have to depend on some principle according to which facts about linguistic sentence meaning carry over to facts about utterance meaning, and such a principle is far from self-evident. There does not seem to be any firm ground for saying anything about the correct interpretation of Mrs. Malaprop's utterance. For our immediate purpose of laughing there does not seem to be any use for any such thing, either. If we do not laugh on account of the contrast between intended meaning and linguistic meaning, we may laugh on account of our imagining that the speaker had actually wanted to say what we are certain that she did not want to say. There is no reason for the hearer to take the meaning which makes her laugh to be the meaning of the utterance.

4.2 Confusion

Confusion is another case which gives philosophers recourse to utterance meaning. Here is a case invented by Perry:

> The man across the street, whom I take to be George W. Bush, is tossing pretzels in the air and catching them in his mouth. I point to him and tell my grandson, "President Bush is tossing pretzels in his mouth. That's very dangerous." I intended to refer to the man across the street by referring to George W. Bush, which I knew I could do by saying "President Bush." The directing intention (now extended to proper names) was to refer to that man I think of when I think "George W. Bush;" the expected upshot was to refer to the man across the street; I planned to refer to the man across the street by referring to President Bush. But I failed; I referred to President Bush and quite possibly said something false about him.
>
> (Perry 2009, p. 198)

The description of the scenario ends with a verdict: the speaker referred to George W. Bush. As Perry describes his case, however, it seems clear that the hearer is presented with evidence in diverging directions. On the one hand, the speaker utters the proper name *President Bush*. On the other hand, the speaker points to the man across the street. The speaker also uses a predicate which aptly applies to the man across the street. I think these circumstances make it very clear that reference here is definitely not a matter of fact or mechanism, but of stipulation or decision. Whether there is reference and, if so, to which object, is something we have to decide upon. But contrary to what Perry suggests, I do not think that any decision comes naturally. The most natural thing for the hearer to do is to ask for the speaker's meaning and to laugh at imagining that George W. Bush actually were tossing pretzels across the street. Any verdict as to what actually was referred to seems uncalled for and is in any case not necessary for laughing.

4.3 Opting for imagined meaning

Imagined meaning may be exploited for a variety of purposes, but in conversational interpretation merriment is certainly the major reason for entertaining a meaning which is neither the speaker's intended meaning nor the most reasonable interpretation of the utterance. Imagined meaning may be carried quite far, I believe. An utterance may serve as an occasion to laugh because the hearer imagines that the speaker had actually wanted to say what she clearly does not want to say or the hearer imagines that she had taken the speaker to say what she clearly does not want to say. But the imagination need not be tied to the actual speaker or hearer at all. The hearer may laugh at what she imagines that someone had wanted to say or at what she imagines that someone had taken the speaker

to say. It is a matter of wit. As such, there are no strictures, yet there are certainly limits. Unless there is a possible and also a quite tangible connection between the imagined meaning and what linguistic meaning and contextual cues present the hearer with, the imagined meaning will appear too far-fetched and as nothing but the hearer's whim. Imagined meaning should not be carried too far if it is to be appreciated. So I believe that imagined meaning, though not at all corresponding to the most reasonable interpretation of the utterance, nevertheless firmly stays within the boundaries of the reasonable.

5 Concluding remarks

I have argued that the debate between intentionalists and anti-intentionalists concerning what constitutes the meaning of an utterance is misguided for two reasons. First, there are no facts about language or communication which have implications for utterance meaning. Second, there is no need for positing or establishing any such thing as the correct interpretation of an utterance. There are indeed circumstances under which a hearer may be inclined to have recourse to the notion of objective utterance meaning, but upon closer inspection she does well without any such notion. If the hearer takes an interest in the speaker's intended meaning, there is no reason for her to pass a verdict on utterance meaning. If the hearer wants to take the speaker to task for her utterance, she is concerned with her own good reasons for taking the utterance the way she did. If the hearer wants to laugh at the utterance, it is imagination which is in play. From the viewpoint of the speaker's and the hearer's interpretive interaction, the notion of utterance meaning is superfluous, or superstitious, as Mrs. Malaprop would put it. However, there is perhaps no glory for me. The options which I have considered may not exhaust the realm of utterance interpretation; the notion of utterance meaning may be necessitated for theoretical reasons which I have not considered. My only hope is to have shifted the burden of proof onto those who think that there is an important use for the notion of the meaning of an utterance.

Acknowledgement

References

Ariel, Mira (2002a): "The Demise of a Unique Concept of Literal Meaning". In: *Journal of Pragmatics* 34(4), pp. 361–402.

Ariel, Mira (2002b): "Privileged Interactional Interpretations". In: *Journal of Pragmatics* 34(8), pp. 1003–1044.

Bach, Kent (1987): *Thought and Reference*. Oxford: Clarendon Press.

Bach, Kent (2017): "Reference, Intention, and Context: Do Demonstratives Really Refer?" In: María de Ponte/Kepa Korta (Eds.): *Reference and Representation in Thought and Language*. Oxford: Oxford University Press, pp. 57–72.

Bach, Kent/Harnish, Robert M. (1979): *Linguistic Communication and Speech Acts*. Cambridge, MA/London: The MIT Press.

Camp, Elisabeth (2008): "Showing, Telling, and Seeing: Metaphor and 'Poetic' Language". In: *The Baltic International Yearbook of Cognition, Logic and Communication* 3, pp. 1–24.

Carroll, Lewis (1872): *Through the Looking-Glass and What Alice Found There*. In: Mark Gardner/Martin Burstein (Eds.): *The Annotated Alice*. New York/London: W.W. Norton, pp. 151–320.

Clark, Herbert H. (1996): *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert H./Wilkes-Gibbs, Deanna (1990): "Referring as a Collaborative Process". In: Philip R.Cohen/Jerry Morgan/Martha E. Pollack (Eds.): *Intentions in Communication*. Cambridge, Mass./London: The MIT Press, pp. 463–493.

Davidson, Donald (1986): "A Nice Derangement of Epitaphs". In: Donald Davidson, *Truth, Language, and History*. Oxford: Clarendon Press, pp. 89–107.

Dodgson, Charles (1896): *Symbolic Logic. Part I: Elementary*, London, Macmillan.

Donnellan, Keith S. (1968): "Putting Humpty Dumpty Together Again". In: *The Philosophical Review* 77(2), pp. 203–215.

Dummett, Michael (1986): "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking". In: Ernest LePore (Ed.): *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell, pp. 459–476.

Durant, Alan (1996): "Allusions and Other 'Innuendo' Meanings in Libel Actions: The Value of Semantic and Pragmatic Evidence". In: *Forensic Linguistics* 3(2), pp. 195–210.

Gauker, Christopher (2008): "Zero Tolerance for Pragmatics". In: *Synthese* 165(3), pp. 359–371.

Gauker, Christopher (2015): "The Illusion of Semantic Reference". In: Andrea Bianchi (Ed.): *On Reference*. Oxford: Oxford University Press, pp. 21–38.

Green, Karen (2001): "Davidson's Derangement: Of the Conceptual Priority of Language". In: *Dialectica* 55(3), pp. 239–258.

Grice, Paul (1967): "Logic and Conversation". In: Paul Grice, *Studies in the Way of Words*. Cambridge, Mass./London: Harvard University Press, pp. 22–40.

Heck Jr., Richard G. (2014): "Semantics and Context-Dependence: Towards a Strawsonian Account". In: Alexis Burgess/Brett Sherman (Eds.): *Metasemantics: New Essays on the Foundations of Meaning*. Oxford: Oxford University Press, pp. 327–364.

Kaplan, David (1978): "Dthat". In: Palle Yourgrau (Ed.): *Demonstratives*. Oxford: Oxford University Press, pp. 11–33.

Kaplan, David (1989): "Afterthoughts". In: Joseph Almog/John Perry/Howard Wettstein (Eds.): *Themes from Kaplan*. New York/Oxford: Oxford University Press, pp. 565–614.

King, Jeffrey C. (2013): "Supplementives, The Coordination Account, and Conflicting Intentions". In: *Philosophical Perspectives* 27(1), pp. 288–311.

King, Jeffrey C. (2014): "Speaker Intentions in Context". In: *Noûs* 48(2), pp. 219–237.

Korta, Kepa/Perry, John (2010): "Intentions to Refer". In: Luca Baptista/Erich Rast (Eds.): *Meaning and Context*. Bern: Peter Lang, pp. 161–186.

Kotátko, Petr (1998): "Two Notions of Utterance Meaning". In: *Proceedings of the Aristotelian Society* 98, pp. 225–239.

MacFarlane, John (2011): "What Is Assertion?" In: Jessica Brown/Herman Cappelen (Eds.): *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, pp. 79–95.

MacKay, Alfred F. (1968): "Mr. Donnellan and Humpty Dumpty on Referring". In: *The Philosophical Review* 77(2), pp. 197–202.

McGinn, Colin (1981): "The Mechanism of Reference". In: *Synthese* 49(2), pp. 157–186.

Montminy, Martin (2010): "Context and Communication: A Defense of Intentionalism". In: *Journal of Pragmatics* 42(11), pp. 2910–2918.

Mount, Allyson (2008): "Intentions, Gestures, and Salience in Ordinary and Deferred Demonstrative Reference". In: *Mind and Language* 23(2), pp. 145–164.

Neale, Stephen (1992): "Paul Grice and the Philosophy of Language". In: *Linguistics and Philosophy* 15(5), pp. 509–559.

Neale, Stephen (2007): "On Location". In: Michael O'Rourke/Corey Washington (Eds.): *Situating Semantics: Essays on the Philosophy of John Perry*. Cambridge, MA: The MIT Press, pp. 251–393.

Neale, Stephen (2013): "Determinations of Meaning." Unpublished manuscript, CUNY/NYU Mind & Language Seminar, February 2013. Available at: http://neale.commons.gc.cuny.edu/files/2012/09/Neale_Determinations_Final.pdf.

Neale, Stephen (2016): "Silent Reference". In: Gary Ostertag (Ed.): *Meanings and Other Things: Themes from the Work of Stephen Schiffer*. Oxford: Oxford University Press, pp. 229–342.

Nunberg, Geoffrey (1995): "Transfers of Meaning". In: James Pustejovsky/Branimir Boguraev (Eds.): *Lexical Semantics: The Problem of Polysemy*. Oxford: Clarendon Press, pp. 109–132.

Pagin, Peter (2002): "Rule-following, Compositionality and the Normativity of Meaning". In: Dag Prawitz (Ed.): *Meaning and Interpretation*, Stockholm, Kungl. Vitterhets historie och antikvitets akad., pp. 151–181.

Perry, John (2009): "Directing Intentions". In: Joseph Almog/Paolo Leonardi (Eds.): *The Philosophy of David Kaplan*. New York: Oxford University Press, pp. 187–201.

Predelli, Stefano (1998): "I Am Not Here Now". In: *Analysis* 58(2), pp. 107–115.

Quinn, Frances (2015): *Law for Journalists*. Boston: Pearson.

Recanati, François (2005): "Literalism and Contextualism: Some Varieties". In: Gerhard Preyer/Georg Peter (Eds.): *Contextualism in Philosophy: Knowledge, Meaning, and Truth*. Oxford: Clarendon Press, pp. 171–196.

Reimer, Marga (1991): "Demonstratives, Demonstrations, and Demonstrata". In: *Philosophical Studies* 63(2), pp. 187–202.

Reimer, Marga (2004): "What Malapropisms Mean: A Reply to Donald Davidson". In: *Erkenntnis* 60(3), pp. 317–334.

Robertson, Geoffrey/ Nicol, Andrew (2002): *Media Law*. London: Sweet and Maxwell.

Saul, Jennifer M. (2002): "Speaker Meaning, What is Said, and What is Implicated". In: *Noûs* 36(2), pp. 228–248.

Sheridan, Richard Brinsley (1775): *The Rivals*. In: Cecil Price (Ed.): *The Dramatic Works of Richard Brinsley Sheridan*, vol. 1. Oxford: Clarendon Press, pp. 35–148.

Shuy, Roger W. (2009): *The Language of Defamation Cases*. Oxford: Oxford University Press.

Smartt, Ursula (2006): *Media Law for Journalists*. London: Sage.

Smit, J.P. (2012): "Why Bare Demonstratives Need Not Semantically Refer". In: *Canadian Journal of Philosophy* 42(1), pp. 43–66.

Speaks, Jeff (2016): "The Role of Speaker and Hearer in the Character of Demonstratives". In: *Mind* 125(498), pp. 301–339.

Speaks, Jeff (2017): "A Puzzle about Demonstratives and Semantic Competence". In: *Philosophical Studies* 174(3), pp. 709–734.

Stokke, Andreas (2010): "Intention-Sensitive Semantics". In: *Synthese* 175(3), pp. 383–404.

Travis, Charles (1989): *The Uses of Sense: Wittgenstein's Philosophy of Language*. Oxford: Oxford University Press.

Wettstein, Howard K. (1984): "How to Bridge the Gap between Meaning and Reference". In: *Synthese* 58(1), pp. 63–84.

Åkerman, Jonas (2010): "Communication and Indexical Reference". In: *Philosophical Studies* 149(3), pp. 355–366.

Åkerman, Jonas (2015): "The Communication Desideratum and Theories of Indexical Reference". In: *Mind & Language* 30(4), pp. 474–499.

Chi-Hé Elder
Negotiating What Is Said in the Face of Miscommunication

Abstract: In post-Gricean pragmatics, communication is said to be successful when a hearer recovers a speaker's intended message. On this assumption, proposals for 'what is said' – the semantic, propositional meaning of a speaker's utterance – are typically centred around the content the speaker aimed to communicate. However, these proposals tend not to account for the fact that speakers can be deliberately vague, leaving no clear proposition to be recovered, or that a speaker can accept a hearer's misconstrual even though the speaker didn't intend it. In such cases, identifying 'what is said' is more contentious, even though communication is arguably no less successful. Building on recent interactionist approaches to meaning, this chapter offers a proposal for 'what is said' that can account for cases of 'imperfect' communication in which the recovery of a speaker's intention goes awry. This involves taking 'what is said' as the co-constructed meaning that the speaker and hearer jointly agree on, giving 'grounded meanings' – those mutually and manifestly accepted into the discourse by speaker and hearer – higher precedence in a theory of communication than simply the speaker's intended meaning. In such cases, rather than speaker and hearer converging at the level of explicit utterance content, communication is expected to be successful as long as the hearer recovers a proposition which is compatible with the speaker's higher-order discursive goal. The hearer's uptake is then allowed to override the content of the speaker's intended meaning in the process of negotiating 'what is said'.[1]

Keywords: background assumptions, co-construction of meaning, hearer inferences, miscommunication, semantic contextualism, speaker intentions, theory of communication, truth-compatible inferences, utterance meaning, what is said

1 Introduction

Following Grice's (1989) seminal writings on 'meaning', the latter parts of the 20th century, extending into the first decade of the 21st century, saw a surge in proposals for what should constitute the unit of analysis for semantic theory: 'what is said'. It was these proposals that were credited at the core of the debate between semantic minimalism and semantic contextualism, what has been termed the so-called 'border wars' (Horn 2006). At the heart of this debate is the question of the extent to which context should play a role in determining 'what is said'. At one end, semantic minimalists purport that the aim of semantics is to determine the invariant properties of language in order that stable truth conditions can be ascribed to abstract sentences, while contextualists, at the other, are concerned with finding the intuitive truth conditions that interlocutors use in everyday communication, thus acknowledging and hence allowing that aspects of context will inform the truth conditions of utterances used in context.[2]

Insofar as a theory of *communication* aims to describe and explain the ways in which people use and understand language in everyday conversation, this chapter puts forward a view on 'what is said' that falls squarely in the contextualist camp. However, this chapter calls into dispute one central assumption that dominant extant contextualist proposals rest on, namely, that successful communication is dependent on hearers recovering speakers' intended meanings.[3] We know that the process of recovering speakers' intended meanings can be more or less straightforward depending on how determinately the meaning can be recovered from the explicitly uttered sentence form. At the more syntactic/semantic end, indexicals and other context-dependent expressions require contextual resolution for their referents; at the more pragmatic end hearers very often engage in complex inferential work to figure out a speaker's intended meaning, taking into consideration an array of contextual information in the process. In short,

---

2 There are also proposals by which 'what is said' does not map onto truth-conditional content, such as Bach's (1994) 'propositional radicals' which pertain to locutionary content, as opposed to *illocutionary* content. I continue in the tradition of equating 'what is said' with the content of illocutionary acts of, e.g. stating and asserting.

3 Here I sidestep the many recent debates on the nature of 'intentions', using 'speaker intentions' to refer to a Gricean notion of a speaker's (reflexively held) intention to communicate something to a hearer, by way of the hearer's recognition of that intention. See Navarro (2017) for an overview of the different uses of the term 'intention' in the philosophical pragmatics literature.

language vastly underdetermines meaning, and, given that communication occurs more or less successfully more or less most of the time, the goal of semantic – and pragmatic – theory is to explain what these meanings are and where they come from. This chapter shows that the problem of recovering speakers' intentions goes further than simply identifying speakers' intended meanings from underdeterminate forms. Rather, one of the central aims of this chapter is to highlight that the problem becomes more complicated when we acknowledge that (a) speakers do not always communicate determinate propositions, and (b) hearers do not always recover propositions that align with those intended by the speaker. While occurrences of (a) are well acknowledged by Relevance Theorists (since Sperber & Wilson 1995), instances of (b) are typically considered to lie outside the remit of a normative theory of meaning, in that such a theory aims to provide generalisations of what interlocutors *should* do, not what interlocutors *actually* do. However, work in interactional pragmatics (e.g. Haugh 2012) provides evidence that interlocutors do have observably patternable ways of resolving misunderstandings, opening the way for a normative theory to account for such recurring patterns of interaction.

Opposed to a view on 'what is said' that relies on speakers' intentions perfectly aligning with inferred content, this chapter builds on recent approaches to propositional meanings informed by interactional pragmatics (e.g. Elder & Haugh 2018) to offer an alternative proposal for the unit of analysis for semantic and pragmatic theory that can account for 'imperfect' communication: cases in which speakers' intentions do not necessarily align with the inferred content (perhaps because there is no determinate proposition to be recognised), and yet communication is arguably no less successful. The idea is that as long as a hearer recovers a proposition which is *compatible with the speaker's overall communicative goal*, the hearer's uptake can contribute to a process of meaning recovery, rendering 'what is said' not as determined solely by the speaker's intentions, nor by the hearer's recovered meaning, but by the *negotiation of meaning between the speaker and the hearer*. The resulting proposition is one that the speaker and hearer co-construct which, in turn, takes on the status of 'what is said'.

This view on 'what is said' is contentious for a number of reasons. First, in some cases it will not necessarily correspond to the speaker's explicit utterance meaning and so the extent to which such a meaning can constitute what is *said* is clearly up for dispute. Of course, the idea that 'what is said' need not be constrained to the logical form (or developments or modulations thereof) but can, on occasion, pertain to an implicature that diverges from the logical form altogether is not new. This latter idea is reminiscent of Jaszczolt's (2005) 'primary meaning', which is a representation of the main meaning that is intended by a Model

Speaker and is recovered by a Model Addressee, or Ariel's (2002) 'privileged interactional interpretation', the speaker's most relevant contribution to the discourse.[4] But where the notion of a co-constructed 'what is said' diverges from that of a 'primary meaning', is for those cases in which the speaker's main intended meaning or most relevant contribution is *not* recovered. In those cases, the hearer's uptake is also allowed to *override* the content of the speaker's primary intended proposition, and instead to point at a proposition which both the speaker and hearer agree on, which, in turn, is compatible with the speaker's overall communicative goal recovered by the hearer. The benefit of opting for such a unit of semantic investigation is that – from the point of view of the speaker-hearer dyad – the meaning that is 'operationalised' (cf Arundale 2013) is the one that is mutually agreed upon by the participants in the process of *communication.*

With this brief overview in place, the structure of the chapter is as follows. In Section 2, I outline some background on the unit of meaning analysis in extant semantic/pragmatic theory. Section 3 presents some evidence to motivate extending the notion of 'what is said' to account for cases of 'imperfect communication'. Section 4 offers an alternative proposal that builds on extant contextualist notions of 'what is said' but that also accounts for potential mismatches, and Section 5 points towards directions for further research.

2 Background: From Grice to Radical Contextualism

Grice's seminal contribution on non-natural meaning, or 'meaning$_{NN}$', set the stage for the received view on the semantics-pragmatics boundary. For him, 'A meant$_{NN}$ something by x' as long as "A uttered x with the intention of inducing a belief by means of the recognition of this intention" (Grice 1957, p. 220), explicitly defining utterance meaning in terms of the speaker's intended meaning, and implicitly defining 'successful communication' in terms of a hearer inferring that intention. Grice acknowledged that sometimes the speaker's intended meaning may depart from the logical form of the utterance as a conversational implicature, in which case meaning$_{NN}$ is composed of 'what is said' – the unit to which truth conditions are applied – and speaker meaning, or 'what is implicated'. On his

---

4 Note that Ariel (2002, 2016) retains a view on 'what is said' that pertains to a context-dependent 'explicature' as per the Relevance Theoretic view (cf. Sperber & Wilson 1995).

view, 'what is said' comprises "the particular meanings of the elements of *S* [the sentence], their order, and their syntactical character" (Grice 1969, p. 87), rendering *sentences*, in abstraction from context, as the input to semantic theory.

Now, while Grice's early pragmatic theory gave us a two-tiered picture of meaning in which the study of semantics is a separate enterprise from that of pragmatics, Grice (1978) himself recognised that sentences underdetermine semantics in cases of ambiguity and reference assignment, and that context is necessarily required for their resolution. However, in making that move, he admitted that some sentences only acquire their 'meanings' once they are uttered in context, paving the way for the future 'boundary dispute' concerning the proper unit of semantic analysis. Since then, a variety of proposals have emerged that aim to offer a conception of 'what is said' that can output intuitive truth conditions of uttered sentences as natural language users use and understand them (e.g. Sperber & Wilson 1995; Carston 2002; Recanati 2004, 2010), alongside a backlash of 'minimalist' views that aim to retain a purely sentence-driven conception of semantics that is devoid of contextual intrusion (e.g. Borg 2004, 2012; Cappelen & Lepore 2005).

Contra the semantic minimalist view is the view which prioritises 'what is said' as a reflection of the ways in which natural language users understand meanings in ordinary discourse. Post-Gricean developments with such a 'contextualist' orientation are guided by the overarching principles that 'what is said' need not be constrained by the sentence form, and that context is allowed to play a significant role in determining the truth-conditional unit. It is this recourse to context that satisfies the aim of generating a semantic unit that aligns with speakers' intuitions about truth conditions. However, different accounts diverge in the extent to which they allow context to influence and inform the unit of semantic investigation. At the more conservative end of the contextualist spectrum is 'indexicalism', where any contextual information that contributes to establishing truth-conditional content has to be attributed to context-sensitive 'slots' in the logical form (e.g. Stanley 2000). On this view, pragmatics is only called upon when the linguistic system demands it, and when the linguistic system demands it, pragmatics is required. This has the aim of ensuring that truth conditions align with speakers' intuitions while ensuring that enrichments are systematic and non-arbitrary. However, when aiming for a conception of 'what is said' that is "in conformity to the intuitions shared by those who fully understand the utterance – typically the speaker and the hearer, in a normal conversational setting" (Recanati 2004, p. 14), we may favour the view that not all pragmatic processes that contribute to semantic content are mandated by covert variables, but that in ad-

dition to these 'bottom-up processes', we may also allow optional 'top-down processes' operating on the words uttered to produce an enriched, 'developed' (e.g. Carston 2002) or 'modulated' (e.g. Recanati 2010) logical form. Thus, we move away from a sentence-centric view on semantics, and towards one that takes *utterances* as the input to semantic theory.

These contextualist options take us some way towards producing intuitive truth conditions of natural language utterances. But, as extensively noted by Jaszczolt (2005, 2010, 2016), sometimes the main proposition expressed goes beyond enriching the extant logical form of the utterance or modulating senses of individual words. In a move to understand 'occasion meaning', Jaszczolt (2016, p. 8) argues that a theory of meaning should be concerned with "the full, intuitively most plausible, meaning as intended by language users on a particular occasion." Note that as we get closer to occasionalism, the possibility of generalisation becomes more difficult; if meaning is constrained by use, we cannot abstract meanings from, e.g. an independent lexicon or grammar, and all we can do is describe the meanings that individual instances of language use give rise to. To get around this problem of the occasionalist, Default Semantics marries an occasion-based conception of meaning with a formalisation by relativising acts of communication to *model communicators*, offering "a formal account of how a Model Speaker constructs meaning in his/her head, and how a Model Addressee recovers this intended message" (Jaszczolt 2016, p. 10). This takes the onus away from *actual* speakers and hearers; the concern is not with actual interpretations of meanings on specific occasions, but with identifying the general mechanisms in meaning comprehension.

Default Semantics, then, is a radical version of semantic contextualism in which the truth-conditional unit pertains to the *primary, intended content* of the utterance. To generate such a unit, the logical form of the utterance may be enriched and, in some cases, even overridden. For example, while the mother's utterance in (1) would pertain to the enriched version (1') on standard contextualist analyses, Default Semantics assumes that the most salient, plausible candidate that the interlocutors are likely to recover is actually the implicit meaning given in (1'').

(1) Child: Can I go skiing?
     Mother: You are too young.
(1') The child is too young to go skiing.
(1'') The child cannot go skiing. (adapted from Jaszczolt 2010, p. 195)

On this theory, word meaning and sentence structure – one source of information in this theory – are not given any priority in determining 'what is said'; rather, in different contexts of utterance, different sources may take precedence over others. So, while in traditional Gricean pragmatics we have two layers of meaning, namely 'what is said' and 'what is implicated' corresponding to 'literal' explicit content versus 'non-literal' implicit content respectively, in this radical version of semantic contextualism, the explicit/implicit divide is supplemented with the primary/secondary meaning divide, where primary meanings (instead of explicit meanings) take on the status of 'what is said'. In the same way that primary meanings can pertain to either explicit or implicit content, secondary meanings too can correspond to explicit or implicit content. The unit of truth-conditional analysis comes about as a result of contextual information, and word meanings interact with other sources of information to generate intended meanings. The result is a completely pragmatic, contextually-driven theory of communication that attempts to reflect cognitive reality.

The unit of analysis of Default Semantics is grounded in Grice's notion of meaning$_{NN}$ insofar as it relies on the assumption that there is a single proposition that a speaker intends to communicate as primary, and that successful communication rests on a hearer recovering that proposition. And indeed, in the ideal case, that is, when speakers' intended meanings are successfully communicated to hearers, the notion of a primary meaning works well: it captures the intuitively most plausible meaning that the speaker intended and that the hearer recovered. However, while sympathetic with prioritising primary meanings when speakers' intended meanings are automatically recovered, this chapter questions what happens to our theory of communication when we move away from Gricean assumptions, examining cases in which (a) speakers don't communicate determinate propositions to be recognised, or (b) hearers fail to recognise speakers' main intended meanings, as failing to explain the process of meaning recovery of such instances would reduce the remit of our theory of meaning to an implausibly constrained, idealised set of communicative interactions. Admittedly, while a normative theory with explanatory power necessarily has to posit a level of idealisation, the fact that we can observe recurring patterns of meaning negotiation provides some rationale for extending the scope of a formal theory of meaning. Such a theory would allow that even 'model' speakers engage in 'imperfect' communication while accounting for the ways in which they successfully resolve such imperfections. The following section thus calls into question the assumption that even model speakers and hearers are expected to communicate according to the 'speaker implicates, hearer infers' mantra. It then moves to propose a

view on 'what is said' that can account for 'imperfect' communication that considers both speaker and hearer perspectives in the negotiation of meaning, thus foregrounding the importance of 'grounded meanings' – those mutually and manifestly accepted into the discourse by speaker and hearer (Clark and Brennan 1991) – in the notion of 'what is said'.

3 What is said when speakers and hearers diverge in understanding

The motivation for a revised view on 'what is said' comes in part from an observation that Grice himself recognised vis-à-vis the calculability of conversational implicatures:

> [...] since there may be various possible explanations, *a list of which may be open*, the conversational implicature in such cases will be an *open disjunction* of such specific explanations, and if the list of these is open, the implicatum will have just the kind of indeterminacy that many actual implicata do in fact seem to possess.
>
> (Grice 1975, pp. 39–40, my emphasis)

The fact that speaker meaning is potentially indeterminate with respect to the precise proposition expressed was a motivation of Relevance Theory (Sperber and Wilson 1995) to take *hearer inferences* as the object of study in place of speaker meaning. Sperber and Wilson (2015) have recently revived this line of argument, demonstrating that expressed propositions lie on a continuum from more to less determinate, with fully propositional forms at one end, and instances of communication that are not paraphrasable at all at the other. Metaphors and hyperbole are included at this latter end of the spectrum: expressions for which although there exists no unique paraphrase that corresponds to the full import of the speaker's intended meaning, the hearer nevertheless can gauge the relevant 'impression'. That is, the speaker makes manifest an array of propositions that are compatible with the explicit content ('explicature' in their terminology) that give rise to a relevant impression, but no single one of these propositions is, or indeed should be, *the* unique proposition that is attributable to the speaker's intended meaning.

The idea that speakers can communicate indeterminate propositions has received some recent attention with respect to the object of semantic inquiry. Savva (2017) examines instances of subsentential utterances, identifying relatively clear-cut cases of syntactic ellipsis that can be given a unique proposition with

determinate syntactic structure, as well as those cases where the recoverability of the communicated thought depends on an interaction between linguistic and extra-linguistic information, and the corresponding proposition lacks a determinate syntactic form altogether (cf. Stainton 2006). Elder and Savva (2018) use the case of open-ended conditionals to show that this is a ubiquitous practice in everyday conversation:

(2) (Casual conversation about common friends)
  A: And Karen and Ian want to buy her half of the mortgage out, so they'll have too much mortgage.
  B: Yeah... It really is...
  A: I know. With Ian only a tennis coach.
  B: Well even now. I mean, if he has good rates, good bank rates, and he's got a steady job...
  A: That's true.
  (ICE-GB, S1A-036: 035)

The incomplete conditional sentence in (2) can be syntactically completed in a number of different, but equally plausible, ways. Possible completions include: 'he'll be able to pay the mortgage', 'he won't have a problem', 'I think he'll be okay', and many others. But even though the speaker may not necessarily have in mind a single consequent with determinate syntactic structure which would yield a full conditional proposition, admissible completions are expected to be members of a set of *equivalent* or *interchangeable* propositions that overlap in their *pragmatic* implications (Savva 2017, Elder & Savva 2018). So, as long as the hearer recovers a message which is compatible with the speaker's intended impression – comprising the set of interchangeable propositions – communication can be deemed 'successful' insofar as the transmission of information is adequate for the conversation at hand. In this case, the hearer's response, 'that's true', indicates that enough information was provided for the speaker and hearer to come to a joint agreement on the content of the utterance. Indeed, even if the speaker lacked a determinate intention vis-à-vis the precise content of the subsentential conditional utterance, it would be unnecessarily time consuming and inconvenient for the hearer to request clarification (cf. Elder & Beaver 2017).

Unlike the Relevance Theoretic position of modelling hearers' inferences, Savva (2017) proposes an amended version of the unit of analysis from Default Semantics, by which 'what is said' corresponds to an 'informationally basic proposition' that includes only the intersection of information that is communicated

by all of the interchangeable, fully sentential propositions. Unlike semantic minimalism, which might co-opt the term 'basic proposition' for a context-free notion of 'what is said' (e.g. Cappelen & Lepore 2005), Savva's basic proposition is 'contextualist' insofar as its content is dependent on both linguistic and extra-linguistic information available in a given context. In the case of (2), the informationally basic proposition would correspond roughly to 'if he has good bank rates and a steady job, money won't be a problem', with the consequent being contextually supplied by previous utterances in the discussion, and the interlocutors' social and cultural knowledge about having a job and one's presumed ability to pay a mortgage (see Elder & Savva 2018 for example representations in the Default Semantics framework).[5]

While the notion of an informationally basic proposition allows a speaker to communicate a deliberately vague proposition that can take several possible interpretations, it still relies on the hearer recovering a proposition that is compatible with a set of propositions that the speaker *intended* to communicate. That is, while the speaker does not point at a specific proposition with determinate structure, it is assumed that the speaker has an intended impression that is being communicated that gives rise to a set of interchangeable propositions. However, there is a growing wealth of empirical evidence that sometimes hearers recover a meaning that extends beyond the remit of a speaker's intended set of propositions, but rather than constituting instances of 'misunderstanding' that are problematic for communication, the hearer's divergent interpretation contributes to a process of information growth that is beneficial to the communicative exchange. So regardless of whether or not the speaker has a determinate proposition in mind that corresponds to his or her utterance or whether that utterance gives rise to a potentially open disjunction of alternatives, it is still possible for a hearer to recover a message that lies outside the 'intended' set.

What I offer instead is a view on 'what is said' that both aims at an empirically-justified view on a propositional, 'primary meaning' that takes both speaker and hearer perspectives into consideration, but that is also grounded in a view in which speakers' intentions and hearers' inferred meanings align. So rather than pursue a view on 'what is said' that is determined either by the speaker's intended utterance meaning or the hearer's inferred content, a viable alternative is to track

---

5 The notion of a 'contextualist basic proposition' is useful when identifying meanings that may be acquired by overhearers and bystanders who are necessarily not involved in the mutual co-construction of meaning, but discussion of the place of such a construct for a theory of communication lies beyond the scope of this paper.

the meanings that interlocutors jointly make operative in the course of a conversation, thereby taking 'what is said' as *the most salient proposition that is interactionally achieved between participants* (cf. Elder & Haugh 2018). Admittedly, sometimes the meaning that is interactionally achieved may diverge significantly from the speaker's initial intended *utterance* meaning, whether that pertains to an explicature or an implicature. In that case, the alignment between speakers' intentions and hearers' inferences is assumed to come at a higher level, namely, at the level of the speaker's *overall communicative goal* which is recovered by the hearer.

The process of interactionally achieving meanings is proposed in formal terms by Elder & Haugh (2018), which highlights that the negotiation of utterance meanings is subject to (at least) a three-part process (see also Haugh 2012). In this three-part process, the meaning of speaker A's utterance is subject to

 i. A's inference about how the utterance will be understood, made available by the utterance itself,
 ii. speaker B's inference about how A expected to be understood, made available through B's response, and, crucially,
iii. A's inference about how B understood A's initial utterance, made available through A's response to B.

Very often, (i) and (ii) align well enough such that (iii) does not do any substantial work in determining the content of A's initial utterance. However, in other cases, a mismatch between (i) and (ii) can be resolved by attending to (iii), which makes available the speaker's inference about how his/her initial utterance is being operationalised while also making manifest where and when divergences in understandings between speakers and hearers occur.[6] So, through this three-turn process, the meaning that is attributed to A's initial utterance can be subject to complex inferential work that is a joint process between the participants, but crucially, those inferences are *formally dependent* on one another, and the resulting proposition is one that cannot (always) be determined by looking at the utterance turn alone. The inference made available by the third turn thus contributes to the co-construction of 'what is said': the proposition that is operationalised by both speaker and hearer.

The theoretical move to take co-constructed meanings as the object of semantic study is inspired by the large and growing body of work in interactional pragmatics (starting from Sacks et al 1974) based on the idea that participants engage

---

6 And sometimes utterance meanings are not settled by the third turn, in which case the negotiation of meaning continues beyond three turns.

in a process of displaying their inferences about previous utterances via their responses to those utterances. However, there is a relative scarcity of works that aim to theorise propositional meanings – in the spirit of post-Gricean pragmatics – in terms of interactional achievement. There is, admittedly, a seeming incompatibility in the epistemological and ontological commitments of interactional analyses versus philosophical semantic analyses of meaning, with the former grounded in a constructionist view of meaning that resides only with the participants themselves, and the latter aiming to represent meanings as somehow objective to the minds in which they reside. However, despite these differences in paradigmatic commitments, as advocated by Haugh (2010), interactional data can be used to inform the minimalism-contextualism debate and, in particular, to test empirically-oriented views on 'what is said' that strive for cognitive reality.

4 Negotiating what is said when understandings diverge

At this point, it is worth giving some examples of how such a co-constructed 'what is said' can arise in conversation. In addition to outlining the process of negotiating 'what is said' through the three-turn procedure, the remainder of this chapter attends to an outstanding question regarding the circumstances under which a hearer's inference is likely to contribute to a process of 'successful communication'; that is, when a hearer's inference will be accepted by the speaker. The simple case is when the hearer recovers a proposition that aligns with the speaker's main intended proposition – even when this intended proposition diverges from the logical form of the uttered content. That is, the hearer infers a meaning that corresponds to either the literal content of the speaker's utterance that the speaker intended to communicate (including modulations or developments thereof, depending on one's contextualist outlook), or a communicated implicature that is both intended by the speaker and derived by the hearer through an interaction of the uttered content with extra-linguistic contextual factors. However, when we add in the considerations that (a) a speaker may not have a determinate proposition in mind that is communicated, and that (b) a speaker may accept a different proposition to the one that was initially intended, the circumstances under which a hearer's inferred content will viably contribute to a process of 'successful communication' is less clear cut.

As discussed, we know from Grice that an utterance can potentially give rise to a number of propositions of which the speaker may or may not be consciously aware, and all of these propositions are available to a hearer to infer. However, as noted by Ariel (2016, p. 23), we must be wary of conflating potential inferences that are *compatible* with what the speaker said with implicatures that are speaker-intended inferences. As she says, 'truth-compatible inferences' may well be speaker *endorsed*, but they need not necessarily be speaker *intended*. Ariel (2016) proposes a battery of tests that distinguish different kinds of pragmatic inferences, including inferences that the speaker intends to communicate directly (including variants of both explicatures and implicatures), but also:

  i. inferences that the speaker intends the hearer to *consider*, but are not intended to be *communicated* by a given utterance ('background assumptions'), and

 ii. inferences that the speaker does not intend the hearer to consider (and are not intended to be communicated), yet that the speaker would be held normatively accountable for and hence should *accept* should the hearer derive them ('truth-compatible inferences').

These latter two kinds of inference will be useful in explaining when and why a given inference may be accepted by a speaker in the negotiation of 'what is said', in the situation where the inferred content does not straightforwardly map onto the speaker's intended content.

To utilise Ariel's terminology, when a speaker produces an utterance, the speaker not only potentially gives rise to a range of propositions that are compatible with the uttered content, but, in addition, makes manifest a set of *background assumptions* that the speaker intends the hearer to consider in order for the hearer to recover a relevant proposition. As Ariel (2016, p. 15) states, background assumptions are entertained by the speaker and are intended for the hearer to consider, but the speaker does not *communicate* them. Background assumptions are akin to Jaszczolt's (2010) 'sources of information', which include world knowledge, stereotypes and presumptions about society and culture, but also even aspects of the situation of discourse that includes knowledge about the participants and the immediate contextual environment. A speaker may expect his or her hearer to draw on particular background assumptions in order to make relevant inferences regarding his or her specific utterance, but it is not in virtue of the *content* of the utterance that these background assumptions are invoked.

Given this apparatus, the proposal of this chapter is as follows: in case a hearer does not recover a proposition that is intended to be communicated by the speaker by a given utterance, communication can still be considered successful

as long as the hearer derives an inference that is not necessarily truth compatible with the explicit content of the utterance and, crucially, not necessarily truth compatible with the speaker's intended meaning, but that is *truth compatible with the intended background assumptions*. In other words, speakers and hearers need not necessarily align at the level of the utterance meaning as it is intended by the speaker at the utterance time. Rather, alignment can occur as long as the relevant background assumptions are accommodated – even when some aspect of the content of the utterance is not.

To exemplify this idea, let us return first to the case of indeterminate propositions as in (2), repeated below.

(2) (Casual conversation about common friends)
    A: And Karen and Ian want to buy her half of the mortgage out, so they'll have too much mortgage.
    B: Yeah… It really is…
    A: I know. With Ian only a tennis coach.
    B: Well even now. I mean, if he has good rates, good bank rates, and he's got a steady job…
    A: That's true.

B's subsentential conditional utterance could arguably be 'completed' to communicate any one of an infinite number of possible propositions. However, despite the fact that the speaker may not have an intended completion in mind with determinate syntactic structure that would give rise to a fully-fledged proposition, there will be a range of *admissible* completions that the speaker would endorse; i.e. there will be a set of propositions that comprises inferences that are truth compatible with the speaker's intended message. The question concerns how to arrive at that set. The idea here is that the speaker's utterance makes manifest a range of background assumptions that the hearer is expected to consider in the process of utterance interpretation: in this case, there is a relevant societal assumption that having a steady job is likely to facilitate an ability to make regular payments on a mortgage. Then, the set of admissible completions of the subsentential conditional becomes restricted to those propositions that are truth compatible with the relevant background assumption. As long as the hearer infers a proposition that is in that set, the speaker is likely to accept the inference derived by the hearer.

The extent to which cases of indeterminate propositions can be considered 'divergent understandings' will depend on one's theoretical commitments, and assuming that a radical contextualist theory aims to recover propositions at the

discursive level as opposed to the level of explicit utterance content, the proposal offered here is expected to be largely uncontroversial. What is more controversial is the attempt to account for cases in which a hearer infers a meaning that does *not* align with the speaker's intended meaning; given that a model hearer has the cognitive power and resources to recover a speaker's intended message, such cases are typically considered beyond the scope of a normative theory of meaning. However, such cases do not always straightforwardly result in communication problems, as the misunderstanding can be resolved through acceptance, with an alternative construal even being preferred by the initial speaker. Such cases are labelled by Clark (1997, p. 589) 'accepted misconstruals', where "speakers present an utterance with one intention in mind, but when an addressee misconstrues it, they change their minds and accept the new construal".

Let us consider Clark's exchange below between himself, a customer (C), and a waiter (W).

(3) W: And what would you like to drink?
     C: Hot tea, please. Uh, English breakfast.
     W: That was Earl Grey?
     C: Right.

Here, W's uptake 'that was Earl Grey?' does not align with C's ostensibly determinate intention to request English breakfast tea. However, rather than correct W, C resolves the misalignment in content by accepting W's construal. As Clark claims,

> I initially intended to be taken as meaning one thing, but I changed my mind. Speakers may accept a misconstrual because they deem it too trivial, disrupting, or embarrassing to correct. Still, once it is grounded, it is taken to be what they mean.
>
> (Clark 1997, p. 589)

So although the speaker may have a determinate intention at the time of producing the utterance, the hearer's response in the second turn may diverge from that initial intention in such a way that is permissible to the speaker. In that case, the speaker accepts the meaning inferred by the hearer in the third turn, which then becomes 'operative' (cf. Arundale 2013) between the speaker and hearer in the current exchange.

Let us take the target utterance that is being negotiated to be C's utterance, 'English Breakfast'. Through this utterance, C makes available an inference that he did, indeed, request English Breakfast tea from the waiter. W's clarificatory

request then makes available her inference on how she thinks C wants to be understood – as requesting *Earl Grey tea* – and, crucially, the third turn by C *confirms* W's inference, thereby operationalising a proposition that was not intended at the time of utterance. However, the fact that both participants were ostensibly satisfied with this interactional outcome, dismissing the interaction as conversation breakdown is counterintuitive. According to the proposal outlined here, the outcome can be explained as follows. First, it is already in the conversational background that C wants to order hot tea from the previous utterance. This information is therefore salient as contributing to the interpretation of the following utterance, 'English Breakfast'. Crucially, it is expected that C would *intend* the hearer to take the information that C is requesting hot tea into consideration when deriving an inference about the meaning of 'English Breakfast'. Now, although the hearer derives an inference that is not truth compatible with C's explicitly *uttered* content, she does derive an inference that is truth compatible with the background assumption that C intends to order hot tea. There are various reasons why C might not correct W in her inference, perhaps because C decides that Earl Grey is preferable to English Breakfast tea, that C does not want to embarrass W, or that C deems it too much effort to correct her. But, as I argue, since an offer of Earl Grey is compatible with the background assumption – and overall communicative goal – of C ordering hot tea, in this case C is happy to accept the misconstrual and retrospectively update the meaning that he ascribes to his initial utterance, landing on a joint view on 'what is said' as a request for Earl Grey.

Note that the post-Gricean may argue that the interaction in (3) *does* display a misunderstanding insofar as the explicit content of the speaker's utterance was not recovered by the hearer. But this is to take a speaker-centric view on the notion of 'what is said'; the contrasting view that prioritises hearer inferences would not aim to represent that misunderstanding as, from the perspective of W, C confirmed W's initial inference and hence no misunderstanding occurred. But the benefit of tracking the meanings that are made jointly operative between the participants in the exchange is that we can incorporate both speaker and hearer perspectives in a single notion of 'what is said', as 'what is said' pertains to a unit co-constructed by participants, attributed to the speaker-hearer dyad.

While I have outlined the proposal for the positive cases, to finish it is worth examining an instance in which communication is not so 'successful', at least insofar as the three-turn process of meaning negotiation does not make manifest a proposition that is operationalised between the participants. Consider the following example (adapted from Varonis & Gass 1985):

(4) [A customer (C) aims to telephone a retail store to buy a new television, but mistakenly calls a representative (R) at a repair centre]

R: Seventeen inch?

C: Okay.

(pause)

R: Well is it portable?

In this extract, the source of trouble comes from the opening utterance 'seventeen inch?' by the service representative, R, about which the speakers struggle to come to an agreed meaning. While on my proposal a successful interaction should involve the customer, C, deriving an inference that is compatible with R's background assumptions that R expects C to consider, it is the divergence in background assumptions about the conversational setting from the outset that lead to troubles in understanding. If there were no such divergence, a viable inference of R's utterance should be compatible with the background assumptions that (i) R works at a repair centre, and hence that (ii) the television under discussion is in C's possession and requires repairing. Thus, a likely truth-compatible completion of the representative R's fragment 'seventeen inch' that R would endorse would likely be 'do you have a seventeen inch television?'. However, the customer C's response 'okay' makes available an inference about the way in which C understood R's utterance, namely as a suggestion – as opposed to a request for information – along the lines of, 'would you like a seventeen inch television?'. That is, C does not derive an inference that is compatible with R's background assumptions, and hence a problem in the interaction becomes manifest; the pause that ensues indicates that R recognises that something is not right with the interaction. Regarding 'what is said', as the two speakers do not operationalise any proposition that is attributed to R's initial utterance, no proposition is grounded, and, from the point of view of the speaker-hearer dyad, nothing is 'said'. It is only later in the interaction, when we would expect the participants to recognise the misunderstanding, and in particular realise their differences in background assumptions, that we would expect the speakers to jointly attribute propositional content to the previous problematic utterance, possibly (jointly) attributing two sets of content to the utterance: one per understanding.

5 Final remarks

In this chapter, I have aimed to show that instances of 'miscommunication' – in the sense of a hearer failing to recognise a speaker's determinate communicative

intention – do not necessarily result in what might be considered communication breakdown. In the ideal case, the meanings that are the most productive at progressing information flow are those that are the main intended meanings of the speaker and that are recovered by the hearer—and these are the cases that have dominated concerns in the field of post-Gricean pragmatics. While some theorists acknowledge that in some contexts and for some sets of interlocutors, the main meaning communicated may depart from the logical form to reflect the main intended speech act, I have aimed to develop the post-Gricean account by taking on board insights from interactional pragmatics, which recognises that information flow isn't always this simple, but it is not always the case that we would want to label an interaction as 'communication breakdown' just because the main message wasn't recovered. That is to say, this proposal is not intended to vitiate extant arguments for speaker- or hearer-centric approaches to representing propositional meaning but is aimed as a supplementary account for when hearers do not straightforwardly recover speakers' intended utterance meanings.

Taking 'what is said' to track the most salient propositions that are mutually made operative between interlocutors is a move towards a theory of communication that accounts for the facts that speakers sometimes leave open the meanings of their utterances and/or that they can, on occasion, allow hearers' interpretations to influence and inform the meanings of their own utterances, and that this is a regularly occurring pattern observable in real-life meaning negotiation. Insofar as speakers and hearers do not always straightforwardly agree on utterance meanings, but that speakers do systematically attend to the three-part architecture of conversational inference, it is not unreasonable to extend a normative theory of utterance meaning in communication such that it can account for and explain such instances. Attending to a three-part process of interaction in which speakers and hearers converge on a joint agreement on 'what is said' is a step forward in this direction that both highlights that the primary utterance meaning of the speaker may not be the one that is inferred by the hearer, and, crucially, where mismatches occur.

While I have exemplified the proposal for a small number of specific instances, the next step will be to test the proposal with more complex interactional data, both exemplifying the three-turn process of meaning negotiation and also showing how interlocutors attend to relevant background assumptions in converging on 'what is said' in interaction. But my broader point is that mismatches in understandings do not automatically result in communication failure; as I argue, a speaker and hearer are likely to come to a joint understanding as long as the hearer infers a content that is compatible with the speaker's background as-

sumptions that the hearer is expected to take into consideration. Thus, the resolution of operationalising 'what is said' may lead to meanings that weren't 'intended' by the speaker at the time of utterance but which are nevertheless compatible with the speaker's higher-order, overall communicative goal.

References

Ariel, Mira (2002): "Privileged Interactional Interpretations". In: *Journal of Pragmatics* 34(8), pp. 1003–1044.

Arial, Mira (2016): "Revisiting the Typology of Pragmatic Interpretations". In: *Intercultural Pragmatics* 13(1), pp. 1–35.

Arundale, Robert (2013): "Conceptualizing 'Interaction' in Interpersonal Pragmatics: Implications for Understanding and Research". In: *Journal of Pragmatics* 58(2013), pp. 12–26.

Bach, Kent (1994): "Conversational Impliciture". In: *Mind and Language* 9(2), pp. 124–162.

Borg, Emma (2004): *Minimal Semantics*. Oxford: Oxford University Press.

Borg, Emma (2012): *Pursuing Meaning*. Oxford: Oxford University Press.

Cappelen, Herman/Lepore, Ernie (2005): *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*. Oxford: Blackwell.

Carston, Robyn (2002): *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.

Clark, Herbert H. (1997): "Dogmas of Understanding". In: *Discourse Processes* 23(3), pp. 567–598.

Clark, Herbert H./Brennan, Susan E. (1991): "Grounding in communication". In: *Perspectives on Socially Shared Cognition* 13, pp. 127–149.

Elder, Chi-Hé/Beaver, David (2017): "The Optimal Rate of Miscommunication". Paper presented at: *Communication & Cognition*, University of Fribourg.

Elder, Chi-Hé/Haugh, Michael (2018): "The Interactional Achievement of Speaker Meaning: Towards a Formal Account of Conversational Inference". In: *Intercultural Pragmatics* 15(5), pp. 593–625.

Elder, Chi-Hé/Savva, Eleni (2018): "Incomplete Conditionals and the Syntax-Pragmatics Interface". In: *Journal of Pragmatics* 138(2018), pp. 48–59.

Grice, Herbert Paul (1957): "Meaning". In: *Philosophical Review* 66(3), pp. 377–388. Reprinted in: Herbert Paul Grice (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 213–223.

Grice, Herbert Paul (1969): "Utterer's Meaning and Intentions". In: *Philosophical Review* 78(2), pp. 147–177. Reprinted in: Herbert Paul Grice (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 86–116

Grice, Herbert Paul (1975): "Logic and Conversation". In: Peter Cole/Jerry L. Morgan (Eds.): *Syntax and Semantics*. Vol. 3. New York: Academic Press, pp. 41–58 Reprinted in: Grice, Herbert Paul (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, pp. 22–40.

Grice, Herbert Paul (1978): "Further Notes on Logic and Conversation". In: Peter Cole (Ed.): *Syntax and Semantics*. Vol. 9. New York: Academic Press, pp. 182–193 Reprinted in: Grice,

Herbert Paul (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 41–57.

Grice, Herbert Paul (1989): *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Haugh, Michael (2010): "Co-constructing What is Said in Interaction". In: Németh T. Enikő/Károly Bibok (Eds.): *The Role of Data at the Semantics-Pragmatics Interface*. Vol. 9. Germany: De Gruyter Mouton, pp. 349–380.

Haugh, Michael (2012): "On Understandings of Intention: A Response to Wedgewood". In: *Intercultural Pragmatics*, 9(2), pp. 161–194.

Horn, Laurence R. (2006): "The Border Wars: A Neo-Gricean Perspective". In: Klaus von Heusinger/Ken Turner (Eds.): *Where Semantics Meets Pragmatics*. Oxford: Elsevier, pp. 21–48.

Jaszczolt, Katarzyna M. (2005): *Default Semantics: Foundations of a Compositional Theory of Acts of Communication*. Oxford: Oxford University Press.

Jaszczolt, Katarzyna M. (2010): "Default Semantics". In: Bernd Heine/Heiko Narrog (Eds.): *The Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press, pp. 193–221.

Jaszczolt, Katarzyna M. (2016): *Meaning in Linguistic Interaction: Semantics, Metasemantics, Philosophy of Language*. Oxford: Oxford University Press.

Navarro, Jesús (2017): "Intention (Including Speech Acts)". In: Anne Barron/Yueguo Gu/Gerard Steen (Eds.): *The Routledge Handbook of Pragmatics*. London/New York: Routledge, pp. 215–226.

Recanati, François (2004): *Literal Meaning*. Cambridge: Cambridge University Press.

Recanati, François (2010): *Truth Conditional Pragmatics*. Oxford: Oxford University Press.

Sacks, Harvey/Schegloff, Emanuel/Jefferson, Gail (1974): "A Simplest Systematics for the Organisation of Turn-Taking for Conversation". In: *Language* 50(4), pp. 696–735.

Savva, Eleni (2017): *Subsentential Speech from a Contextualist Perspective*. PhD thesis, University of Cambridge.

Sperber, Dan/Wilson, Deirdre (1995) [1986]: *Relevance*. Oxford: Blackwell.

Sperber, Dan/Wilson, Deirdre (2015): "Beyond Speaker's Meaning". In: *Croatian Journal of Philosophy* 15(44), pp. 117–149.

Stainton, Robert (2006): *Words and Thoughts: Subsentences, Ellipsis and the Philosophy of Language*. Oxford: Oxford University Press.

Stanley, Jason (2000): "Context and Logical Form". In: *Linguistics and Philosophy* 23(4), pp. 391–434.

Varonis, Evangeline Marlos/Gass, Susan (1985): "Miscommunication in Native/Nonnative Conversation". In: *Language in Society* 14(3), pp. 327–343.

Brian Ball
Playing Games, Following Rules, and Linguistic Activity

Abstract: In this paper I propose that game-playing situations may be regarded as involving *agents* standing in the *playing* relation to abstract *games* with essential rules; and I inquire whether linguistic activity can be so regarded as well. Timothy Williamson (2000) has argued that it can, and that assertion in particular should be thought of as a move in such a game, governed by the rule *assert only what you know*; but Ishani Maitra (2011) argues against the possibility of such an analysis on the basis of what I take to be a refinement of the thought that we follow no such strong, truth-entailing rule. I defend the Williamsonian proposal by rejecting accounts of rule-following that would allow the argument to succeed. Arguably, playing a game does require that agents follow its rules; but agents do, in the relevant (weaker) sense, follow a strong rule of assertion. And even if this account of rule-following should turn out to be mistaken, I suggest that it is possible to play a game while flagrantly violating its rules.

In section 1, I introduce the game playing analysis, Williamson's application of it to linguistic activity through his knowledge rule for assertion, and what I call 'The Rule-Following Argument' against his view. In section 2, I consider two analyses of rule-following, one of which (from intention, not conformity) might be taken to vindicate that argument; but in section 3, I suggest that the adoption of a third (sensitivity) analysis, or an alternative (rule-governance) account of game play, suffice to undermine it. In section 4, I respond to an objection; and in section 5, I engage directly with Maitra's concerns about flagrant rule-violations. I briefly conclude, in section 6, that, despite arguments to the contrary, linguistic activity can be regarded as involving game play.

1 Introduction: Language, Games, and the Rule-Following Argument

Consider your favourite game – say, chess, or soccer.[1] When this game is played, a complex social phenomenon occurs, which we describe by saying that people are playing chess, or soccer. Perhaps we can make progress in understanding such phenomena by taking grammatical appearances at face value. In particular, the grammar suggests that in such circumstances there is a *game* G that people play, and that *playing* is a relation between people, or *agents* more generally, and games.[2]

We have a pre-theoretical understanding of what agents are: rational beings having both preferences and the ability to make choices amongst possible courses of action. This simple account might need to be refined upon further reflection, but it will serve our present purposes. It will be instructive, however, to ask after the natures of the other elements involved, namely games and the playing relation.

What, then, are games? We may take them to be abstract objects individuated by their rules, thought of as essential to them.[3] Although we ordinarily think that a game like soccer can change its rules, on the view urged here this is not (strictly speaking) true. Rather, when something occurs which, as players of games, we would normally describe as a rule change, as theorists we may regard this as a

---

1 As Papineau (2017, p. 278) notes, some but not all games are (constitutive parts of) sports – soccer is, for instance, while chess is not; and equally, some but not all sports involve games (2017, p. 274) – again, soccer does, but running does not. What is essential to sport, on Papineau's view, is that its primary purpose is the exercise of a physical skill (2017, p. 274). This is clearly not essential to games.

2 Arguably, when children entertain themselves by playing, there is no game that they play. But this poses no problem for the analysis proposed in the main text of the transitive use of 'play', since it allows for the possibility that when this verb is used intransitively, the play in question is not a relation, i.e. is not the playing *of* anything: for, being restricted in its scope, it makes no pronouncement at all regarding such cases; in particular, it does not predict that this consists in playing *some game or other*.

3 Note that if the rules of a game are essential to it, then it has those rules necessarily: hence, necessarily, any game identical to it has those rules; and so those rules serve to individuate that game not merely within, but across, possible worlds.

change in social tradition; members of the society under study cease to be interested in playing one game G, and begin to take interest in another, typically very similar game G', which they proceed to call by the old name.[4]

What does it take, for any game G, to play G? An initially plausible answer is that in order to play G it is both necessary and sufficient to follow the rules of G. One of the aims of this paper will be to explore this hypothesis, and with it the nature of rule-following. A second, more fundamental aim will be to investigate whether linguistic phenomena can be profitably regarded as involving games played by conversational participants.

Timothy Williamson (2000) has argued that we should adopt this theoretical perspective: in particular, he has claimed that the speech act of assertion, like a move in a game, is governed by a rule which is essential to it. More specifically, Williamson considers the possibility that assertion is governed by a rule with a simple form, namely, that of the C rule:

> (CR) One must: assert that p only if p has C.

This schema has many instances, but two are of particular interest. The first is the truth rule:

> (TR) One must: assert that p only if p is true.[5]

The second is the knowledge rule:

> (KR) One must: assert that p only if one knows that p.

Williamson argues that assertion is governed by the knowledge rule (KR); but it is important to note that if this is so, it is also governed by the truth rule (TR), for knowledge entails truth.[6]

---

4 See Williamson (2000, p. chapter 11). Note that this is consistent with the letter, though perhaps not the spirit, of Papineau's (2017, p. 258) suggestion that sports are individuated (in part) by their histories, provided we carefully distinguish sports from the games (if any) they involve. That is, perhaps a sport can change its rules (over time) even if the game it involves (at any one time) cannot.

5 Weiner (2005) argues that assertion is governed by the truth rule (TR); it is unclear whether he intends to be taken as holding that assertion is a move in a language game essentially governed by this rule.

6 Williamson also thinks that (KR), unlike (TR), *individuates* assertion – that it is the only speech act uniquely and constitutively governed by the knowledge rule. But this further claim is not pertinent to our present concerns.

Ishani Maitra (2011) rejects the view that assertion is a move in a language game. More precisely, if we call a rule having the form of (CR) "strong" just in case that which replaces "C" entails that p is true – i.e. just in case the norm it imposes on assertion is at least as strong as the truth rule (TR) – she rejects the view that assertion is essentially governed by a strong rule. Her argument (discussed below) can be regarded as a refinement of what I'll call 'The Rule-Following Argument':[7]

  i. Playing, and hence making moves in, a game requires following its rules.
 ii. Speakers often make assertions without following any strong rule. Therefore,
iii. Asserting does not consist of making a move in a game involving such a rule.

This is a powerful argument, and potentially devastating to the view that linguistic activity is rule-governed, and consists of making moves in a language game. But if the argument is to succeed, an account of rule-following must be provided which makes it plausible both that to play a game is to follow its rules, and that assertion does not involve following any strong rule. In the next section I suggest two simple analyses of rule-following, one of which might be thought to make both of these premises true. In the subsequent section I offer a response on behalf of the game-based theorist of language to the argument so construed.

2 Playing games and following rules: Conformity and intention

One extremely simple suggestion is that to follow a rule is simply to conform to it. A second, slightly more complex proposal is that following a rule consists of intending to conform to it. If either of these hypotheses is correct, we can substitute the analysans ('conforming to', or 'intending to conform to') for the analysandum ('following') in the Rule-Following Argument to generate what we might call 'The Conformity Argument' (in the first case) and 'The Intention Argument' (in the second), each having the conclusion that assertion is not a move in a language game that is subject to a strong rule. And, of course, even if the proposed

---

7 Given her conception of rule-following (see below), Maitra would not endorse the Rule-Following Argument itself: but it is inspired by her own argument, together with her citation of Rawls as holding that "[t]o engage in a practice, to perform those actions specified by a practice, means to follow the appropriate norms" (1955, p. 26) or rules; and it provides a useful starting point which will serve to structure the discussion which follows. We will consider Maitra's own argument in some detail below.

analyses should be mistaken, we can consider these arguments in their own right. So we can ask: Is either hypothesis correct? And: Is either argument sound? In what follows the second of these questions will be at the fore, though there will also be some discussion pertinent to the first.

Merely conforming to the rules of G is clearly not sufficient for playing G. One can easily imagine a comedy of errors in which, through a sequence of happy circumstances, someone utterly ignorant of the rules of a game G is treated as a hero by virtue of conforming to the rules of G and having therefore been taken as a winner. Arguably, the comedy (such as it is) results from the discrepancy between appearance and reality: the hero appears to the other characters to be very good at playing the game, when in reality (as discerned by the audience) he is not playing the game at all. The claim that in order to play a game one must intend to conform to its rules, by contrast, predicts this result, and thereby explains the source of the comedy.

Conforming to the rules of G is not necessary for playing G either. Maitra (2011, p. 281) imagines a case in which someone is pitched three strikes, but everyone, including the umpire, loses count, and the batter receives another. One of the rules of baseball is violated in this case; but, Maitra claims, quite plausibly, that the people involved are playing baseball despite this departure from the rules. It follows that the Conformity Argument is not sound: its first premise is false.[8]

Maitra herself concludes from this case that following the rules (or norms) of G is not necessary for playing G on the grounds that "following a norm requires at least conforming to it" (2011, p. 281).[9] She then goes on to argue that, nevertheless, *flagrantly* violating the rules of a game is not compatible with playing that game. And from this, together with the fact that speakers sometimes flagrantly violate strong rules such as the truth rule and the knowledge rule while nonethe-

---

8 Although I have not spelled out this argument, I have given explicit instructions above for how to construct it from the Rule-Following Argument. This makes it determinate which premise is the first and which is the second. I leave it as an exercise to the reader to undertake this routine task. (Similar remarks apply, *mutatis mutandis*, to the Intended Conformity Argument and others.)

9 This is, of course, consistent with the simple suggestion with which we began this section, namely that to follow a rule is to conform to it; however, the fact that Maitra says 'at least' here suggests that she thinks something more is required, perhaps intentionally conforming. This is not a hypothesis regarding rule-following that is considered further here: crucially, it is to be distinguished from the claim that rule-following is intending to conform (which allows for failure to do so).

less making assertions (for instance, when they are obviously lying), she concludes that asserting is not a matter of making a move in a language game involving such strong rules.

Maitra's own account of flagrant rule violation builds in an epistemic component: "a failure to conform to a norm is *flagrant*", she says, "if it is intentional and sufficiently [epistemically] marked" (2011, p. 282, italics original). However, it might be thought that this epistemic component plays a role only in drawing to the attention of other relevant participants that the violator is not playing the game anymore, i.e. is not engaged in cooperative behaviour. This undermines their motivation to cooperate and to play: yet the epistemic aspect of flagrant rule violation merely provides evidence that the original character is not engaged in the cooperative behaviour of playing a game; it does not constitute that failure. If we remove it from Maitra's definition of a flagrant violation of the rules, all that is left is an intentional violation. Arguably, then, Maitra's purposes are well enough served by the suggestion that to follow a rule is simply to intend to conform to it, and that following the rules of a game in this sense is necessary for playing that game. (In any case, we will consider Maitra's proposal on its own terms below.)

Before closing this section, I would like to consider two cases which might be thought to support this hypothesis. First, take the handball committed by Thierry Henry in the World Cup 2010 qualifier against Ireland. In sudden death extra time Henry handled the ball, then passed it to his teammate Gallas, who scored, giving France the victory. Henry intentionally violated the rules of soccer by playing the ball with his hand. Assuming he was rational (so that he did not, in addition, intend to conform to the rules), the current proposal predicts that Henry was *not* playing soccer at the time of the offence,[10] even if he intended to be *taken* (e.g. by the referee) to be playing the game. Indeed, one might say that this is what cheating consists in: intending to be taken to be playing a game while not in fact playing. The fact that the proposal provides a neat analysis of cheating might be seen as an advantage.

The second case is that of the card game Cheat. In Cheat the object is to get rid of one's cards. There are rules which tell us which cards can be disposed of legitimately in which circumstances; but cards are played face down, and it is permissible to lie about which cards one has played – this is what constitutes the "cheating" which gives the game its name. If one is challenged one must then

---

10 Considering a related claim, Papineau says, "This is the kind of thing that gives philosophy a bad name" (2017, p. 61). It is therefore worth noting that I am not endorsing it; in fact, I will go on to reject the view that entails it.

show the cards one has played; and if one has lied, one must collect all of the cards played in that round. Teenage experiments have revealed that if one deliberately violates *this* rule, the game quickly degenerates, and no one wants to play. Despite its name, it seems that one does not play Cheat by cheating.

Prima facie, then, it is plausible that playing a game requires following its rules in the sense of intending (*de rebus*)[11] to conform to them. In the next section, however, I argue that this is not the case, thereby defending the game-playing analysis of linguistic phenomena against the Intention Argument. (The Conformity Argument, recall, has already been dealt with.)

3 Alternative approaches: Sensitivity and Rule-Governance

It is clear that strong rules such as the truth rule (TR) and the knowledge rule (KR) are intentionally violated by liars, who nevertheless make assertions; thus, if following a rule consists of intending to conform to it, asserters do not invariably follow any strong rule. If Maitra's conclusion that making an assertion does not consist of making a move in a language game is to be resisted, the advocate of the game-theoretic approach to the study of linguistic activity must deny that rule-following, in this sense, is necessary for playing a game – that is, s/he must reject the first premise of the Intention Argument.

One way to do this is to accept that playing a game requires following its rules, but to deny that following rules consists of intending to conform to them; that is, one might reject the second (intention to conform) analysis of rule-following suggested in the last section. Arguably, Williamson favours this approach: "Constitutive rules" such as the knowledge rule (KR),[12] he says, "do not lay down necessary conditions for performing the constituted act" (2000, p. 240); and what follows reveals that he means that *conforming* to the rule is not necessary for performing the act. "Nevertheless," he continues, "some sensitivity to the difference – in both oneself and others – between conforming to the rule and breaking

---

11 If someone unacquainted with either baseball or cricket, though having heard a little about them, were to encounter people playing cricket, she might conjecture that they were playing baseball. If she were to observe for some time, learning the rules, she might join in; yet despite her *de dicto* intention to conform to the rules of baseball, she would be playing cricket.

12 It is common, following Rawls (1955), to distinguish two kinds of rules: thus, Searle (1969) distinguished constitutive from regulative rules, as did Lewis (1979). Williamson (2000) is alluding to this tradition in speaking of constitutive rules. See below for further discussion.

it presumably is a necessary condition of playing the game... or performing the speech act" (2000, p. 240). Those who display the requisite sensitivity might be said to be *following* the rule. On this approach, following a rule consists in having some kind of knowledge, perhaps tacit, that certain acts, or moves in a game, are governed by, or subject to, that rule, and employing it to guide one's actions.[13] According to advocates of this 'Sensitivity' approach, while the first premise of the Intention Argument is false, it is the second premise of the Rule-Following Argument that is mistaken.

A second possible response is simply to deny that following the rules of a game G is, in any sense, necessary for playing G. According to Don Ross (2011), for instance, game theorists in other domains typically assume that any situation in which "at least one agent can only act to maximize his utility through anticipating... the responses to his actions by one or more other agents is... a game" (2011, p. section 2.2); or better, since games themselves are abstract objects "partly defined by the payoffs assigned to the players" (2011, p. section 2.7, italics original) in the various outcomes of the game, this constitutes a *game-playing situation*. It seems obvious that conversations are game-playing situations in this sense: what one gets out of asserting something, for instance, depends upon whether one's audience believes what one says. Moreover, one can be in a game-playing situation so construed, and so be playing a game in this sense, without intending to conform to the rules of this game. "Game theory has been fruitfully applied in evolutionary biology," Ross notes, "where species and/or genes are treated as players" (2011, p. section 6) despite not having any intentions at all; it suffices that these "players" have payoffs (defined in terms of reproductive success) as described by the abstract game structure.[14] Finally, it seems that one can make sense of the thought that the moves these players make are subject to, or governed by, norms, or rules: if a given move (e.g. refusing to cooperate with the police in the prisoner's dilemma, and instead choosing to cooperate with the other player of the game) can be shown not to yield the best outcome for a given player on the assumption that other "players" do what yields the best outcome for them, it can be regarded as incorrect; and players ought not to make such moves. This holds even if the players do not, because they cannot, *intend* to conform to a rule prohibiting such play. Thus, it is necessary, if one plays a game G,

---

13 One challenge for those pursuing this strategy will be to spell out the notion of tacit knowledge in such a way that it is plausible that following the rules of a game G in this sense is not only necessary, but also sufficient, for playing G. This will not be attempted here.
14 Here we can no longer regard players as agents in the sense of the introduction.

only that one's actions are *governed* by the rules of G, not that one *intend to conform* to these rules. On this 'Rule-Governance' approach, it is the first premise that fails in both the Intention, and the Rule-Following, Argument.

It might be objected, however, that playing what we ordinarily think of as a game – a cultural artefact of a certain sort – requires something different than playing what game theorists think of as a game. The reason is that these "artificial" games involve rules which do not seem to be reducible to what it is otherwise rational for an agent to do given what s/he can anticipate other rational agents will do. For instance, why should Henry not play the ball with his hands except that it is against the rules? On the face of it, we do not seem to be able to explain the fact that this act is impermissible by appeal to antecedent distributions of payoffs over possible courses of action as in the "natural" games played, for instance, by genes and species. But if assertion is governed by a strong, truth-entailing C rule, it must be a move in an artificial game; and so performing it, our objector says, will require more than having the relevant payoffs.

But game theory can explain the playing of what I have been calling artificial games: it is because, due to past interaction, people expect each other to tell the truth, or refrain from playing the ball with their hands, that (if they are rational) they penalize those who don't; accordingly, the payoff for an agent in a situation in which he lies or commits a handball but is caught are low, and it is rational for him to endeavour to avoid such an outcome. Since this explanation is presented in terms of what is expected, however, it requires treating assertion as a move in a game the playing of which requires following its rules in the sense of having at least tacit knowledge that its moves are governed by those rules – that is, it requires Sensitivity, and not just Rule-Governance.

It remains only to show that the current (Sensitivity) hypothesis regarding game-play can account for the cases which were taken to motivate the alternative (Intention) proposal of the last section. Our comic hero, however, did not have the payoffs associated with winning the game in which he found himself: accordingly, since knowledge is factive, he did not tacitly know his actions to be governed by its rules; from which it follows, on the current proposal, that he was not playing the game. By contrast, Henry seems to have been sensitive, in a pertinent way, to the rules of soccer – for instance, he would likely have accepted a punishment without much protest had it been imposed upon him. His cheating, then, did not consist in his intending to be taken to be playing while not playing; it consisted, rather, in his intentionally violating the rules while knowing that they were in force. Finally, one can't play Cheat by persistently cheating, since in doing so one does not exhibit the required sensitivity to its rules.

I suggest, therefore, that the sense (if any) in which it is necessary, if one is to play a game such as soccer or chess, that one follow its rules, asserters do follow a strong C rule such as the truth rule (TR) or the knowledge rule (KR); accordingly, we may regard linguistic activity as a form of game-play.

4 Regulative and constitutive rules: Clarification, objection, reply

It might be suggested that the proposed account of when an agent counts as playing a game is insufficiently general for my purposes here, and in particular that it was in Henry's interests, in the situation he found himself in, to commit a handball, thereby violating the rule of the game he was playing that prohibits such acts: thus, no good game-theoretic sense has been given to the claim that he ought not to have performed this act; nor has an adequately robust sense in which he was governed by the rules of soccer been articulated.[15] In this section I will respond to this objection, clarifying the above proposal in the process. This will then afford the opportunity, in the next section, to consider the role the epistemic component in Maitra's notion of a flagrant rule-violation might be thought to be playing in her argument, and to discuss an insight on playing games due to David Papineau (2017). I begin by spelling out the general framework of this paper, and the worry, in somewhat more detail.

I have so far been speaking indiscriminately about 'rules'. But distinctions can be made. Thus, some of the rules of a game G tell us what the moves are, and which of them are permissible under what circumstances. Other rules tell us how the score – broadly construed so as to include such information as which inning(s) it is (in baseball or cricket), how much time remains (in soccer or basketball), or whose turn it is (in chess) – changes (in some cases) as a result of the performance of those moves; and ultimately, what it is to win, lose, or draw at G.[16] Rules of the first kind deal principally with the moves of the game: David Lewis (1979) called them *specifications of correct play*; and he called those of the second kind, which are concerned with the game as a whole, *specifications of the kinematics of score*. The rules of these two types will, however, typically be interdependent: which moves are permissible will depend upon what the score is; and

---

15 Thanks to an anonymous referee for raising a version of this objection.
16 See Dummett (1958) for the claim that a specification of a game will include a classification of end states as winning, drawing, or losing ones.

what the score is will depend upon which moves have been made, permissibly or impermissibly. Together they comprise the *constitutive* rules of the game G; and it is these rules which are essential to G.[17]

In addition to the constitutive rules of a game G, however, we may also recognize that there are *regulative* rules which concern it. While the former may be stated as indicatives, the latter most naturally take the form of imperatives.[18] Following Lewis (1979) again, we may take these imperatives to be of two types: *directives concerning score*; and *directives requiring correct play*. In particular, we may say that there is just one directive concerning correct play, namely: *Play correctly!* Similarly, there is only one directive concerning score: *Win!*[19] What makes these rules pertain to G is the fact that they implicitly employ terms defined by the constitutive rules of G. More specifically, we might re-write the regulative rules of G thus: *Make your action correct play in G! Make winning at G the outcome of your action!* These regulative rules of G are not essential to G; rather, they are involved in the *playing* of G. Thus, the suggestion that playing a game involves following its rules can now be rephrased as the hypothesis that it consists in following the regulative rules concerning that game.[20]

Of course, as stated above, Williamson's rules seem to be simultaneously constitutive and regulative; they contain the deontic modal expression "must", which allows them to be formulated in the indicative mood, while at the same time containing (something of) the force of an imperative. But these two components of such rules can be separated from one another; any rule given in the form

---

17 If G had different constitutive rules than it has, it would be a different game – i.e. different from G itself; which, of course, is to say that it couldn't have had different constitutive rules. (See Williamson (2007) for a defence of the claim, relied upon here, that *necessarily p* can be defined as *if p were not the case, p would be the case*.)

18 See Searle (1969) and Lewis (1979).

19 Or perhaps better: *Prefer winning to drawing, and drawing to losing!* See below for some further discussion.

20 Maitra also draws a distinction between constitutive and regulative norms (2011: 280) or rules – though not in the Lewisian manner pursued here. She also proposes a third category of norms that are 'connected to purposes that players have *qua* players of the game' (2011, p. 278). She says, "The purpose of a player, at least *qua* player, is to win. If he does not attempt to win at all, then he is not playing the game" (2011, p. 278). So, the directive concerning score (Win!), which I regard as a regulative rule of any given game, Maitra takes as 'purposive norm' (2011, p. 278). But Maitra's claim that one must try to win in order to count as playing a game is plainly false – otherwise a tennis player could not fix a match, nor could a boxer throw a fight. (My view is not subject to this criticism since I have rejected the claim that one must intend to conform to the regulative rules of G in order to play G.)

of the C rule can be rewritten as a constitutive indicative rule (where G is the language game to which assertion belongs):

(C-CON) An assertion that p is correct in G only if p has C.[21]

Together with the directive concerning correct play, *make your action correct in G*, this will be equivalent to the original C rule. The above account of game-play in general can therefore be applied to the idea that assertion is a move in a language game that is subject to a C rule.

We are now in a position to state the above objection more precisely. As we have seen, game theorists think that agents can play games without intending to conform to their rules: it suffices if they are in some sense able to perform the various moves available in the game, and have the relevant payoffs. They certainly need not conform to the regulative rule, or directive, concerning *score*: that is, they need not win, maximize their payoff, or more generally do what they ought in this regard. Moreover, some "agents" – those that have no minds – can't even intend to do so. Indeed, they don't even need to act as if they are doing so – that's just doing what they ought to do. (For instance, in biology, some genes or species die out, and this fact might be explained by the fact that they are playing some game, but playing it badly, i.e. not as they ought.) Still, they are subject to the norm, rule, or directive concerning score – they ought to conform. But this does not, as yet, give any way of understanding the idea that such agents are subject to the directives concerning *correct play*. In fact, it seems to preclude the possibility, since (it might be thought that) in order to be playing a game the agent must have available precisely the moves specified in the game; but, in order to be subject to a requirement to conform to the rules, it must be possible, but not permissible, for them to violate those rules. Without the distinction between the permissible and the possible, we can't make sense of a normative requirement. Indeed, the problem deepens: we haven't even given a sense to the idea that games have rules of correct play, let alone to the idea that agents are governed by them.

Nevertheless, it seems there is a way to articulate a notion of a rule of play in (indefinitely) repeated games; to use this to capture a distinction between possible and permissible moves in such games; and to articulate a sense in which agents ought to conform to such rules. Consider repeated prisoner's dilemma. There are two moves available to each player in any given interaction: cooperate

---

21 Notice that this provides a response to Maitra's complaint that 'neither the knowledge norm nor the truth norm has the right *form* to be a constitutive norm' 2011, p. 282). The schema (C-CON) clearly has the right form.

with the other player, or defect. There are also strategies that players might follow: for instance, an agent might play tit-for-tat, cooperating in the first interaction, and then doing what the other player did in the previous round. But then, we can, in effect, parse this as involving a rule and a punishment: the rule is to cooperate; the punishment (for violating the rule on a given occasion) is one round of defection. Alternatively, an agent might play the grim strategy: cooperate until your partner defects; thereafter always defect (in interactions with this partner). Here the rule is the same (cooperate), but the punishment is different (defection evermore, not just for one round). If the punishment for rule violation is great enough, players will be subject to the rule of correct play (cooperate) because of long term payoff while nevertheless having the option of violating it in order to obtain short term payoff. Indeed, if there is epistemic uncertainty, or metaphysical indeterminacy, about how many rounds will be played, there might be genuine uncertainty, or even indeterminacy, about what is the best thing to do – conform to the rule or violate it – in order to maximize utility, even though agents are subject to the rule in the sense that violation will trigger punishment.

This, I believe, offers a response to the objection articulated above: for it makes sense of the idea of a rule of correct play in game-theoretic terms; and it allows for a sense in which there are moves available that are impermissible – in particular, moves that constitute rule-violations and so will trigger punishments – even if they maximize payoff.

5 Flagrant rule-violations: Maitra's argument and Papineau's proposal

The above discussion suggests a role for Maitra's idea of an epistemically marked, or flagrant, rule-violation in relation to the playing of games. Obviously, a violation of a rule (such as the rule that one must cooperate in repeated prisoner's dilemma) cannot be punished unless it is noticed. And this suggests that a flagrant rule-violation is one that is sufficiently epistemically marked to be noticed. But noticed by whom? The agent(s) responsible for issuing the punishment, of course.

In many games, such as soccer, baseball, or competition chess, that agent will be the referee, umpire, or arbiter. Indeed, in the example considered by Maitra immediately prior to her definition of a flagrant rule-violation, the umpire plays a prominent role. Maitra asks us to consider a case in which Derek

has earned three strikes in a plate appearance. As is usual, the umpire has ruled that Derek is out. However, Derek refuses to leave the batter's box. He admits that he has already earned three strikes, but insists that the game would be more enjoyable for the spectators if he were allowed an extra strike.... [I]f Derek is sufficiently recalcitrant about being allowed the fourth strike, we are inclined to say that he has stopped playing baseball altogether.

(Maitra 2011, p. 281)

Maitra's argument, it seems, can be regarded as the No Flagrancy Argument, which results from replacing 'following' in the Rule-Following Argument with 'not flagrantly violating'. A further example, however, shows that even the No Flagrancy Argument is not sound, and that its first premise in particular is false. In the same 2010 World Cup in which Henry committed his handball against Ireland in the group stages, Uruguayan striker Luis Suarez used his hands to stop what would have been a certain goal for his quarter-final opponent, Ghana, in the last minutes of the match. The referee saw the infraction, issued a penalty shot, and had Suarez sent off. But Ghana failed to score, and eventually lost; Uruguay progressed to the semi-final. Suarez' handball could not have been more flagrant. It was clearly intentional, and everyone saw it, including the referee. Yet despite this, Suarez was playing soccer. In fact, he was trying to win.

Papineau has a proposal regarding the playing of games that readily accommodates this case. According to Papineau (2017, p. 67), one is playing a game (e.g. of soccer) just so long as one respects the authority of the referee. Since Suarez respected the decision of the referee to send him off, he counted as playing the game. (The account also accommodates the case of Henry since, as was noted above, he would likely have accepted any punishment given by the referee had he been caught. Indeed, in some ways, Henry's handball was more egregious than Suarez's, since its being advantageous to him to commit it required that he not be caught, whereas this was not so for Suarez: the former required a covert rule-violation; the latter was happy with a flagrant one.) Moreover, Papineau's proposal vindicates the verdict that the first premise of Maitra's No Flagrancy Argument is false.

Nevertheless, there are games which have no (external) referee and to which, at least on the face of it, Papineau's account is inapplicable. In language games, and e.g. the repeated prisoner's dilemma, it is the players themselves who issue the punishments. (In the case of Cheat, considered above, the same people act both as players and as referees; and it is when, in their role as referees, they fail to uphold the rules that the game degenerates, and the other people involved lose interest in playing.) Still, in these games, as we have seen, the players may be subject to rules of correct play (such as the rule requiring cooperation in the re-

peated prisoner's dilemma), so long as others issue punishments when they detect violations – which, of course, requires that those violations be sufficiently epistemically marked. Nevertheless, agents can be said to be playing such games when their actions are guided by the knowledge that they are subject to such rules, even in the weak sense that (sufficiently marked) violations will trigger punishments.

Can this account of game-play accommodate flagrant rule-violations? I think the answer is yes. Consider the case of (then) White House Press Secretary Sean Spicer's claim, concerning the crowd at President Trump's inauguration that "this was the largest audience to ever witness an inauguration, period." Photographic evidence showed otherwise. Spicer's claim therefore involved a flagrant violation of strong norms of assertion such as the truth and knowledge rules – though, of course, it was an assertion nonetheless. This fact can be accommodated, however, on the assumption that the White House that issued this statement did so in full awareness that it was in violation of these rules. Arguably, the central aim of this assertion was to draw to the attention of those who might be concerned about its mendacity that the White House was not worried about any punishments they might issue in response. In other words, it was arguably a signal that the President considered his position to be sufficiently strong that he need not conform to the rules. But this message could not be (intentionally) delivered unless it was (at least tacitly) known that assertions are subject to a truth-entailing norm – the action had to be chosen in a manner that was sensitive to the fact that assertion is governed by such a rule.[22]

It might nonetheless be argued that this kind of sensitivity to rules is not sufficient for following them. Perhaps not. I am not primarily concerned here with providing an analysis of rule-following, but with defending the view that assertion can be regarded as a move in a language game that is subject to a strong rule. As Maitra's No Flagrancy Argument makes clear, what matters for these purposes is to show that one can play a game while flagrantly violating its rules; and the original Rule-Following Argument can be resisted, provided simply that there is no single notion of rule-following which makes both of its premises true simultaneously.

---

22 Perhaps the claim made here is overly strong. The key point, for present purposes, is that the advocate of the game playing analysis of linguistic activity can explain this flagrant rule-violation on the assumption that asserting is making a move that is governed by a strong rule. Perhaps some alternative explanation is also available. But the existence of this one suffices to show that the No Flagrancy Argument fails. (Thanks to an anonymous referee for pressing this point.)

6 Conclusion

I have defended the possibility of treating linguistic activity as a kind of game-play. In the process I have argued that, in order to play a game, it is not necessary that one conform, or intend to conform, to its rules. I have also suggested that to follow a rule one must have some tacit knowledge of it, and that rule-following so construed is involved in the playing of language games; but even if this account of rule-following should prove to be incorrect, it is nevertheless the case that flagrant rule-violations are compatible with game-play.

References

Dummett, Michael (1958): "Truth". In: *Proceedings of the Aristotelian Society* 59, pp. 141–162.

Lewis, David (1979): "Scorekeeping in a Language Game". In: *Journal of Philosophical Logic* 8(1), pp. 339–359.

Maitra, Ishani (2011): "Assertion, Norms, and Games". In: Jessica Brown/Herman Cappelen (Eds.): *Assertion: New Philosophical Essays*. Oxford: Oxford University Press, pp. 277–296.

Papineau, David (2017): *Knowing the Score*. London: Constable.

Rawls, John (1955): "Two Concepts of Rules". In: *The Philosophical Review* 64(1), pp. 3–32.

Ross, Don (2011): "Game Theory". In: Edward Zalta (Ed.): *The Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/archives/fall2011/entries/game-theory/

Searle, John (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Weiner, Matthew (2005): "Must We Know What We Say?" In: *The Philosophical Review* 114(2), pp. 227–251.

Williamson, Timothy (2000): *Knowledge and Its Limits*. Oxford: Oxford University Press.

Williamson, Timothy (2007): *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.

Joanna Odrowąż-Sypniewska

Subsentential Speech Acts, the Argument from Connectivity, and Situated Contextualism

Abstract: The most commonly given examples of subsentential speech acts are expressions such as "Nice dress", "From Spain", "Where?", etc., uttered in such circumstances that the speakers uttering them are regarded as "making moves in a language game", e.g., stating, asking, promising, etc. The argument from connectivity is one of the most important arguments for the claim that such utterances – contrary to appearances – are in fact ellipses, i.e. sentential speech acts. The argument uses examples from inflectional languages, such as Polish or German, in which allegedly subsentential speech acts (e.g. "Obiema rękami" (Both hands. INS), said by a father to his little daughter drinking chocolate milk from a glass) appear in cases other than the nominative. Those who think that they are just fragments of longer unpronounced sentences have no problem in explaining where the case comes from, but the answer is more problematic for those who think that such utterances are truly subsentential. In this chapter, I argue that this argument is by no means conclusive and that the defenders of subsentential speech acts need not be worried by connectivity effects. I suggest a situated contextualist account of such acts, which is based on Recanati's moderate relativism (see Recanati 2008), on which connectivity can be explained. However, *pace* Recanati and following Perry (1986), I argue that, at least in the case of subsentential speech acts, we have to postulate unarticulated constituents in explicit contents as well as in situations of evaluation. The explanation of connectivity effects appeals to Perry's (1986) idea of the utterance completing the thought whose other constituents are already in a given situation. Speakers use cases other than the nominative in order to simplify the process of enrichment for the hearers. The cases make it easier to determine which completion of the articulated content is the intended one.

Keywords: connectivity, ellipsis, limited ellipsis, moderate relativism, situated contextualism, subsentential speech acts

1 Introduction

Subsentential speech acts are subsentential utterances by means of which one "makes moves in a language game" (Dummett 1973, p. 194), i.a., asserts, requests, asks, etc. Among the examples given by Robert Stainton are the following:

1. "Sanjay and Silvia are loading up a van. Silvia is looking for a missing table leg. Sanjay says, 'On the stoop'" (Stainton 2006, p. 5).
2. "Benigno gets into a taxi and says 'To Segovia. To the jail'" (Stainton 2006, p. 5).
3. John enters the pub and says: "Three pints of lager" (see Stainton 2006, p. 158).
4. "Meera is putting jam on her toast. As she scoops out the jam, she says 'Chunks of strawberries'. Anita nods, and says 'Rob's mom'" (Stainton 2006, p. 115).
5. "A father is worried that his daughter will spill her chocolate milk. The glass is very full, and she is quite young, and prone to accidents. He says, 'Both hands'" (Stainton 2006, p. 5).

The traditional view has it that in order to do something with an expression the speaker must embed it in a larger syntactic structure so that what he utters is a complete sentence. For instance, Michael Dummett writes:

> A sentence is, as we have said, the smallest unit of language with which a linguistic act can be accomplished, with which a 'move can be made in the language game': so you cannot do anything with a word – cannot effect any conventional (linguistic) act by uttering it – save by uttering some sentence containing that word
>
> (Dummett 1973, p. 194).

On the face of it, the conversational situations described by Stainton are clear counterexamples to Dummett's thesis. In each of them, what the speaker utters is a subsentential expression, and yet we have the intuition that he or she achieves something. Sanjay suggests that the missing table leg is on the stoop, Benigno requests that the taxi driver take him to the jail in Segovia, John asks for three pints of lager, Meera notices that the jam contains chunks of strawberries, while Anita states that it was made by Rob's mom, and finally the father tells his daughter that she should hold the glass with both hands. Among philosophers working on the topic of subsentential speech acts, there are those who think that such examples can be reconciled with the traditional view and those who think that the examples demonstrate that the traditional view is false.

2 Saving the traditional view

2.1 Subsentential utterances are not speech acts

There seem to be two options that one can take to save the traditional view. Namely, either one can deny that in the above situations any speech acts have been performed, or one can argue that the expressions uttered are not genuinely subsentential. One may, for instance, claim that an utterance, in order to count as a speech act, must have determinate content and determinate illocutionary force, and since subsentential utterances do not fulfil this condition they cannot be regarded as full-blooded speech acts. For instance, Jason Stanley considers the case in which "a thirsty man (...) staggers up to a street vendor and utters: water" (Stanley 2000, p. 407). Stanley argues that the man hasn't performed a speech act. Firstly, his utterance has no determinate illocutionary force – it is unclear whether it was meant as an assertion, a request, or a command. Secondly, it has no determinate content – even if we assume that its force is assertoric, it is still not clear whether he says that he wants water or that the vendor should give him water. The problem with this view is that although there are many nonsentential utterances which are too indeterminate in content and force to count as speech acts, saying that all such utterances are not determinate enough would result in denying many *sentential* utterances the right to be speech acts. Elugardo and Stainton give the following example (Elugardo and Stainton 2004, pp. 465–466): Maria, who is both a supervisor and a good friend of Susan's, says to her: "You must turn in your final report before you leave in the afternoon". The utterance has quite determinate content, but it is very likely that the force of this utterance is indeterminate. It is somewhere between an assertion, an order, and a request.[1] Nevertheless, we do not want to argue that Maria didn't do anything as a result of this indeterminacy; we don't want to say that she didn't perform any speech act at all. Similarly, if a client says to a bartender "Lager, please", it may be indeterminate whether he meant that he would like to get a lager or that he would like the bartender to pull him a lager, but it would be overly pedantic and unrealistic to argue that no speech act has been performed.[2] Thus, one should not

---

[1] Elugardo and Stainton stress that it is not merely epistemologically indeterminate, but because of Maria's professional and personal relationship with Susan, it may also be metaphysically indeterminate (see Elugardo 2004, p. 466).

[2] However, the fact that such speech acts have indeterminate content may be used as an argument against their being ellipses. See below.

conclude from the observation that a given utterance's content and force are not fully determinate that this utterance does not constitute a speech act.

Moreover, even if this strategy is convincing for some cases of subsentential utterances, it would be much harder to apply it to all of them.[3]

2.2 Subsentential utterances are not genuinely subsentential

2.2.1 Extralinguistic context raising linguistic expressions to salience

The second way to save the traditional view is to argue that the utterances in question are not genuinely subsentential. The advocates of this view claim that Sanjay's utterance "On the stoop" is similar to John's answering "Bill" to the question "Who bought the tickets?". John utters "Bill", but his utterance is elliptical, and what he really says is "Bill bought the tickets".[4] Similarly, when Sanjay seems to say "On the stoop" he really says "It is on the stoop" and his utterance is a case of ellipsis. One of the main proponents of this view is Jason Merchant, who claims that most cases of subsentential utterances should be regarded as ellipses. Elugardo and Stainton notice that there are many similarities between common examples of ellipses and subsentential speech acts. For instance, speakers mean more than what they pronounce and what is conveyed is not merely conversationally implicated. Like ellipses, subsentential utterances have complex meaning properties (e.g. they can be ambiguous and stand in entailment relations). They are also subject to grammatical constraints (such as connectivity effects – see below) (See Elugardo and Stainton 2005, pp. 5–6). Nevertheless, Elugardo and Stainton point out that there is also one crucial difference: "subsentences (…) seem to be more easily licensed by features of the world" (Elugardo and Stainton 2005, p. 6). The important distinction between Sanjay's utterance

---

3 Stanley is well aware of that. He proposes a divide-and-conquer strategy according to which allegedly subsentential utterances can be divided into three groups. Whereas one such group contains examples that do not qualify as speech acts, the examples in the other two groups are either ellipses with linguistic antecedents determined by extralinguistic contexts or shorthands (see Stanley 2000).

4 Merchant argues that we can see that this is a case of ellipsis, because the answer 'inherits' the syntactic structure from the question. Thus, one cannot, for instance, answer this question with "By Bill" even though he could have answered "The tickets have been bought by Bill" (see Merchant 2010, p. 157). However, on the contrary, Progovac (2006) indicates that there are correct short answers that cannot easily be expanded to full sentences (e.g. Who ate the pie? Me).

and John's utterance is that only the latter has a linguistic antecedent. John's answer 'inherits' the particular syntactic structure of the question it addresses. Conversely, it is a characteristic feature of subsentential speech acts that they often occur in discourse-initial positions with no linguistic antecedents. When Sanjay says "On the stoop" he does not answer any articulated question. His utterance is the first in that context and it seems that there is nothing from which his utterance could 'inherit' the missing structure. Two answers to this problem can be found in the literature. Stanley claims that:

> explicitly providing a linguistic antecedent by mentioning it is only the simplest way to provide it. There are other methods of raising linguistic expressions to salience in a conversation without explicitly using them
>
> (Stanley 2000, p. 404).

He argues that if extralinguistic context is rich and determinate enough it may raise linguistic expressions to salience. Consider the following example:

> suppose that a group of friends, including John and Bill, has gone bungee jumping. Every member of the group is watching Bill, who is the first to muster the courage to bungee jump. As Bill is standing eight stories above the water on the platform of a crane, ready to plummet into the water below, Sarah, aware of John's terror of heights, turns to one of the other friends and utters (4) [John won't], shaking her head
>
> (Stanley 2000, p. 405).

Stanley admits that Sarah's utterance is perfectly felicitous, but claims that, despite appearances, it does not occur in a discourse-initial position, since the expression "bungee jump" has been made salient by the extralinguistic context and can serve as the linguistic antecedent.

This strategy is controversial and has been met with criticism. For instance, Stainton argues that although the extralinguistic context can make a certain activity salient, it cannot make salient a linguistic expression. Even in the situation described above, although it is pretty clear that people are focusing on bungee-jumping, there is more than one way in which Sarah's utterance could be completed ("John won't jump", "John won't do it", in addition to "John won't bungee jump"). I'll argue in section 4.1 that someone who wants Sarah's utterance to be regarded as a common ellipsis cannot allow for such variation.

2.2.2 The limited ellipsis hypothesis

Another strategy has been proposed by Merchant. He puts forward a limited ellipsis hypothesis for linguistically discourse-initial cases, according to which

"the context makes manifest either a salient action, licensing ellipsis of *do it* (...), or a salient entity, licensing ellipsis of a deictic (...)" (Merchant 2004, p. 727). Thus, if you consider a situation described by Stanley:

> Bill walks into a room in which a woman in the corner is attracting an undue amount of attention. Turning quizzically to John, he arches his eyebrow and gestures towards the woman. John replies: (5) A world famous topologist
>
> (Stanley 2000, p. 404).

the limited ellipsis hypothesis will allow you to say that (5) is elliptical for "This is a world famous topologist". Similarly, if Sanjay says to Silvia, who is looking for a missing table leg, "On the stoop", we may say that his utterance is elliptical for "It is on the stoop". On the other hand, if Alison and John look into a room, see a horrible mess on the floor, and Alison says "Lauren", then Alison's utterance is elliptical for "Lauren did it" (see Stainton 2006b, p. 99). Stainton notices that the context makes the horrible mess salient, but presumably it must be the making of the mess that is salient for the limited ellipsis hypothesis to work. The hypothesis is the *limited* ellipsis hypothesis because there are only two possibilities (either "it is/this is" or "do it"). This allows Merchant to argue that the process involved in the interpretation of such utterances is mere disambiguation. No mind reading or 'guess-timating' is needed to recover the structure of the complete sentence. Thus, the autonomy of the language faculty is not threatened.

Merchant himself acknowledges that his hypothesis doesn't work for all cases.[5] One might try to use it in our fourth case: one might argue that Meera actually says "These are chunks of strawberries", while Anita answers "Rob's mom did it (i.e. the jam)",[6] but it cannot help with cases 2, 3, and 5. For some problematic cases, Mer-

---

5 Limited ellipsis hypothesis lists two exceptions (namely "it/this/that is" and "do it") to the general rule that ellipses need linguistic antecedents. Stainton notices that the hypothesis could work for more cases if we "increase the inventory of exceptions" (Stainton 2006b, p. 108). One could argue that not only "did it" and "this is" can be elided without the linguistic antecedent but also, e.g., "give me", "take me" etc. However, if we do this, then the autonomy of the language faculty becomes threatened. The hearer, in order to recover the structure, has to rely heavily on extralinguistic context and guess what the speaker had in mind. See Stainton (2006b, p. 109).

6 Actually, both Stainton and Merchant doubt that the limited hypothesis helps this case. Stainton argues that Meera communicates "This jam contains chunks of strawberries" (see Stainton 2006a, p. 115). Obviously, the limited ellipsis hypothesis does not provide such content. However, one might argue that what Meera *says* is "These are chunks of strawberries", "This jam contains chunks of strawberries" is merely implicated.

chant suggests semantic ellipsis analysis and scripts. According to semantic ellipsis analysis, the semantic values of some nonsentential expressions contain free variables whose values are to be determined by context. Thus, for instance, "on the stoop" contains an unbound variable whose value is determined by a relevant assignment function. Merchant claims that "by having the semantic 'slots' in the meaning (semantic value) of the phrase uttered, they can all be type <t>, propositional" (Merchant 2010, p. 177). The idea of scripts comes from Schlank and Abelson (1977). In everyday exchanges, speakers follow certain scripts. They know that they do so, and they expect their interlocutors to do the same. The mutual knowledge of the script allows the speakers to anticipate future utterances. It also results in certain phrases being expected in a given context. So, for instance, when Benigno gets into a taxi and says "To Segovia", he will be understood as saying "Take me to Segovia", thanks to the script of giving instructions to a taxi driver. If Beningo uttered the same phrase in a pub while speaking to the bartender, he probably wouldn't be understood. Merchant claims that highly conventionalized contexts make particular linguistic expressions salient, "though of course there may be some indeterminacy in exactly which verb has been conventionally made manifest" (Merchant 2004, p. 731).

Now, it seems to me that admitting that which verb has been made manifest may be indeterminate gives the game away. Such an expression cannot be regarded as an ellipsis (see below). Moreover, the example with the father saying "Both hands" still remains problematic, since there is no script waiting to be used in this situation.

3 Rejecting the traditional view

If one maintains that subsentential utterances may actually perform speech acts and yet is not convinced by arguments establishing them as covertly elliptical, one has no choice but to reject the traditional view. In doing so, subsentential utterances are taken to be both genuine speech acts and, crucially, *genuinely* subsentential. Thus for instance, Kent Bach writes:

> Examples like these [i.e. like 1–5 – J.O.-S.] seem to show that the above assumption [i.e. the assumption that one needs a sentence to perform a speech act – J.O.-S.] is false, so obviously so that it might better be regarded as a pretense or, if you like, an idealization
>
> (Bach 2008, p. 739).

The proponents of this view include, among others, Ellen Barton, Eros Corazza, Robert Elugardo, Alison Hall, and Robert Stainton.[7] The most extensively developed standpoint of this kind has been proposed by Stainton, who devoted several papers and a full monograph to its defense. He argues that "[t]he propositional content of sub-sentential speech acts is arrived at by grasping (a) a content from language and (b) a content from elsewhere, which is never 'translated into' natural language format" (Stainton 2006a, p. 156).

Stainton does not claim that subsententials express propositions. He acknowledges that they are sub-propositional, but nevertheless argues that one can perform a speech act, for instance an assertion, by using a subsentential. He distinguishes two cases. Either the speaker utters a phrase whose content is an argument for which the content provides the propositional function, or he utters a phrase whose content is a propositional function while the context provides the argument for that function (Stainton 2006a, p. 156). Such utterances have asserted, as well as merely conveyed, content. The asserted content is "that proposition which results from minimally adding to the content of the bare phrase actually uttered so as to arrive at a proposition" (Stainton 2006a, p. 60).

Stainton writes:

> My view is this. What drives the hearer to look for the missing argument or function is (a) the non-propositional meaning of the item uttered, a matter of type semantics, together with (b) something about pragmatics. Specifically, what does (part of) the work is the recognition that the speaker intends to communicate a proposition – so she cannot mean$_{NN}$ what her words encode, since they don't encode a proposition. In contrast, the hearer realizes, her interlocutor could mean such-and-such proposition – got from combining the language-derived content with a salient entity of some kind. And so on, in familiar Gricean fashion
>
> (Stainton 2006a, p. 158).

The problem with this explanation lies in its reliance on Gricean mechanisms to determine what is asserted. As is well known, Grice didn't write about what is asserted but rather about what is said. Moreover, his remarks concerning what is said are not particularly developed. He wants what is said to be propositional, intended by the speaker, and close to conventional meaning. As far as the context-sensitivity of what is said is concerned, Grice allows only disambiguation and reference resolution for indexicals. What is important here is that the Gricean notion of what is said is not inferentially-arrived-at content. The latter is regarded as otherwise communicated content, e.g. conversational implicature. The hearer

---

7 See, e.g., Barton (1990), Corazza (2011), Elugardo and Stainton (2004), Hall (2009).

realizes that what the speaker said cannot be the intended meaning, and hence tries to infer the implicature instead. Therefore, it seems to me that Stainton's mention of "familiar Gricean fashion" is illegitimate in this context, for he appeals to Gricean inferential mechanisms to explain how the asserted content is arrived at.

4 Problems for the extant views

The views that claim that subsentential speech acts are genuinely subsentential are called pragmatics-oriented, whereas views that argue that such acts are in fact sentential are called semantics-oriented. In my opinion, the best argument against the semantics-oriented views is the argument from indeterminacy,[8] whereas the best argument against the pragmatics-oriented views is the argument from connectivity. On the ellipsis view, subsentential utterances are fragments of sentences, so if the shape of those sentences is indeterminate, the view loses much of its plausibility. As we will see, some of the proponents of this standpoint claim that indeterminacy can be accommodated, but I do not find their arguments convincing. On the other hand, on the pragmatics-oriented view subsentential utterances are *not* fragments of larger syntactic structures, so for the proponents of this view it is a challenge to explain why certain connectivity effects occur.

4.1 Against the semantics-oriented view: The argument from indeterminacy

As Elugardo and Stainton notice, in order to support the ellipsis view it is not enough to argue that the speaker could have uttered a complete sentence instead of a sentence fragment. Nor is it enough to claim that "the speaker intended the

---

8 Other important arguments against the ellipsis view are the arguments from ambiguity and embedding. Stainton points out that on such views the same sound corresponds sometimes to a word or phrase and sometimes to a full sentence. Thus "Lauren" at some times may be used as a proper name, whereas at other times, as we have seen, it may be a sentence. Stainton doesn't consider such ambiguity a major drawback, but it seems to me that it is a serious problem for the view – especially given that the posited ambiguity does not stop there. "Lauren" used as a sentence is still ambiguous between "Lauren did it" and "This is Lauren". Another problem that Stainton mentions is that expressions which are elliptical sentences should embed just like ordinary sentences. However, while "I couldn't eat the worm, but John did it" is well-formed, "I couldn't eat the worm, but John" is not (Stainton 2006b, pp. 104–105).

hearer to find some sentence or other, not actually uttered, and to fill in *its* slots to arrive at what is asserted" (Elugardo and Stainton 2004, p. 454). What matters is whether the complete sentence has been processed both by the speaker and the hearer.[9] The argument from indeterminacy is just an observation that when a (not very polite) customer in a café says:

> Black coffee!

His request is indeterminate between (i.a.):
  i. Black coffee, please.
  ii. Can I have a black coffee?
  iii. Could I get a black coffee?
  iv. I'd like a black coffee.
  v. Can you give me a black coffee, please.

All the above are legitimate ways of expanding the customer's order. Moreover, if you asked him whether, for instance, he meant (i) rather than (ii), he would most likely not be able to give a decisive answer. However, if in such a case it is not determinate which of the above is actually meant by the speaker (and recovered by the hearer), then – it seems to me – the ellipsis view is in trouble. What I find surprising is that the proponents of this view, who are aware of the problem, do not take it very seriously. For instance, Martí writes:

> How do we know, in the grammatical approach, whether what the grammar outputs is (...) *I would like a coffee, please!* (followed by movement and ellipsis), as opposed to, say, (...) *I would like to have a coffee, please!*, or *I would like to order a coffee, please!*? We don't. The grammar is not in the business of determining what a speaker wants to say. Its job is to impose constraints on how one must say whatever it is one wants to say
>
> (Martí 2015, p. 441).

And Merchant claims:

> [the elided VP] need not have a determinate propositional content, if by determinate we mean that the hearer can determine precisely what the speaker had in mind
>
> (Merchant 2004, p. 720).

I do agree with Martí, of course, that grammar's job is to "impose constraints on how one must say whatever it is one wants to say"; however, in order to save the

---

9 Merchant does claim that hearers recover complete sentences. For him, inaudible structure is "part of the structure computed in normal language processing" (Merchant 2006, p. 31).

ellipsis view, one needs more than that. To argue that subsentential utterances are in fact sentences whose parts have been elided one must be prepared to provide the full syntactic structure with the allegedly elided fragments. Saying that someone who utters "Black coffee" meant either "Give me a black coffee" or "I'd like a black coffee" just won't do. Such a view is no different from the view which acknowledges that what has been uttered is a subsentential expression used to voice a propositional request. Moreover, as we have seen, it is not only the case that the hearer cannot determine precisely what the speaker had in mind, but also that the speaker himself may not have had any determinate proposition in mind.

4.2 Against the pragmatics-oriented view: The argument from connectivity

The argument from connectivity is based on the observation that in many languages subsentential utterances are subject to grammatical constraints, in particular to case matching. Obviously, this phenomenon is much easier to explain if one assumes that such expressions are mere fragments of larger syntactic structures. Part of the sentence is not pronounced, but it exists and the pronounced fragment has to appear in such a form as to fit into the unpronounced structure. Paradoxically, one of the cases given by Stainton, namely the case featuring the worried father, can be used as an example here. It is one of Stainton's favorite examples because it is really hard to see it as a case of ellipsis. Even Merchant's limited ellipsis hypothesis is of no help here (see Merchant 2010). However, as it turns out, this case is problematic for Stainton in another respect. If we translate it into Polish (which is an inflectional language), then the father might say:

Obiema rękami. (Both hands. INS)

or

Obie ręce.[10] (Both hands. NOM)

A proponent of the ellipsis view may easily explain this by saying that what the father really said is "Trzymaj obiema rękami" (Use both hands). The proponent

---

10 Initially I was convinced that the father would say "Obiema rękami" and not "Obie ręce", but my observations of my husband talking to my 3-year old son proved me wrong.

of the pragmatics-oriented view has a much more difficult task, for there is no equally easy explanation forthcoming.

One can find numerous examples along the same lines. For instance, in Polish a customer ordering black coffee will say

> Czarną kawę. (Black coffee. ACC)

rather than

> Czarna kawa. (Black coffee. NOM).

Stainton realizes the importance of this objection, but he tries to neutralize it by providing examples in which there is no case matching. He describes the following situation (which is slightly modified and discussed in detail by Merchant):

> Hans and Franz are playing a boring game one day in which each person takes turns naming an object which reminds him of a particular person. Their conversation consists of sentences such as
> (50) Die Lampe erinnert mich an meinen Onkel Wolfram. (…)
> 'The lamp reminds me of my uncle Wolfram.'
> They go their separate ways and a few days later, Hans is sitting in a bar when Franz walks in the door. Hans points at a nearby beer-stained old wooden table and says,
> (51) Mein Vater!
> my. NOM father ('My father!')
> (Merchant 2010; see Stainton 2006, p. 107).

Here Hans uses the phrase in the nominative even though case matching would dictate the form "meinen Vater". The example is quite elaborate, but it is easy to describe a less complex situation. For instance, one of Stainton's other cases may be used here if we translate it into Polish. The case is as follows:

> after two weeks of cold and rainy weather in mid-summer, in a part of Canada that is usually hot and sunny, Brenda ran into Stan. Brenda looked up at the sky and said 'Nova Scotia'
> (Stainton 2006, p. 6).

Stainton claims that the uttered phrase's content is an argument for which the context supplies the function "the weather here is similar to _____". If this is so, then in Polish the function is "pogoda tutaj jest podobna do pogody w_____" which requires the argument in the locative: "Nowej Szkocji (LOC)". Nevertheless, Brenda could not say "Nowej Szkocji"; the only form that she could use is "Nowa Szkocja" (Nova Scotia. NOM). An even simpler scenario involves the utterance "Nice dress". Stanley writes that it is short for "This is a nice dress" (Stanley 2000, p. 409), whereas Hall says that it communicates "You're wearing a nice dress" (Hall 2009,

p. 237). In Polish "Ładna sukienka" (Nice dress. NOM) can be used to mean either "This is a nice dress" or "You're wearing a nice dress". However, only in the former sentence "ładna sukienka" occurs in the nominative ("To jest ładna sukienka"), whereas in the latter "Masz ładną sukienkę", "ładna sukienka" is in the accusative. Hence, again, the grammatical case of the subsentential utterance does not match that of the relevant fragment.

Thus, the proponent of the pragmatics-oriented view may argue that we have reached a stalemate. The cases in which there is no case matching counterbalance those in which there is case matching. However, I think that we may do better than this. In what follows, I'll describe a version of the pragmatics-oriented view which is based on Recanati's moderate relativism, and I will argue that on such a view we may explain the phenomenon of case matching.

5 Situated contextualism and the argument from connectivity

The account I propose is based on Recanati's framework of moderate relativism.[11] Recanati's relativism presupposes two principles: duality and distribution. The principle of duality says that to assign a truth-value to a proposition we need both a content and a circumstance of evaluation, whereas the principle of distribution says that the determinants of truth-value distribute over content and circumstance. To those two principles, Recanati adds the principle of economy which says that the richer the content, the poorer the circumstances can be, and vice versa (Recanati 2007, p. 33–34). Thus, for instance, the content of "It's raining here and now" uttered on October 20th, 2017 in Warsaw will be that it is raining on 20.10.2017 in Warsaw and the circumstance will consist exclusively of the corresponding possible world, whereas the content of "It's raining" uttered in the same situation will be that it is raining, and not only the world but also the day and the place will belong to the circumstances. Perry (1986) distinguishes articulated and unarticulated constituents of the content (both are the things that the proposition is *about*) and the things that the proposition *concerns* (i.e. those that belong to the circumstances of evaluation). In that text, he does not use the term "constituent" for the things that the proposition concerns. However, in his later papers and in particular in *Critical Pragmatics* (co-authored with Kepa Korta) he

---

11 This section is based on my "Subsentential speech acts: a situated contextualist account" (forthcoming).

defines unarticulated constituents as objects to which a speaker refers without using a corresponding referring expression (see Perry and Korta 2007, p. 102). Therefore, in what follows we will assume that unarticulated constituents may belong either to the content or to the circumstances. Thus, unarticulated constituents – in general – can be divided into those that the proposition is about (unarticulated constituents of the content) and those that the proposition concerns (unarticulated constituents belonging to the circumstances). According to Perry, there are unarticulated constituents of both sorts, whereas Recanati argues at length (2007, part 9) that there are no unarticulated constituents in the content. According to Recanati, all unarticulated constituents are things that the proposition concerns and hence belong to the circumstances of evaluation.

In my view, Recanati's moderate relativism is an ideal framework for analysing subsentential speech acts, albeit only if his objection to unarticulated constituents in the content of subsentential expressions is omitted.[12] Subsentential speech acts might be seen as a limiting case for his principle of economy: their content is extremely impoverished and the circumstances must be quite rich for them to qualify as speech acts. The framework allows a two-level evaluation: we may argue that "From Spain" as uttered by John holding a letter is not a sentence, hence it has no propositional content and no truth-value. Nevertheless, seen as an utterance it can be assessed as true or false in a particular discourse situation. The required elements from the discourse situation are added to its explicit content, so as to make it truth-evaluable. However, as I have already mentioned, it seems to me that at least in the case of subsentential speech acts we have to postulate unarticulated constituents in their contents. I agree with Perry and Recanati that we understand utterances like "It's raining" and "It's three o'clock" without even thinking about the place and time of the utterance and about time zones respectively, because the situation of the utterance provides all the relevant parameters; we do not need to consider them. Therefore, the proposition does not have to be *about* them and they do not have to be included in the content. Whether or not the speaker thinks about them, they are determined by the situation, and the hearer will know how to determine the value of those parameters if the need arises. However, the case of subsentential utterances is different. It doesn't suffice that the missing elements are in the situation: they have to be cognitively articulated (at the level of thought). If John utters "From Spain" hold-

---

12 In *Perspectival Thought*, there is an example of subsentential speech ("Very handsome!" used as a comment on someone's appearance), but as far as I am aware Recanati has never discussed subsentential speech acts in any detail.

ing a letter, but his utterance is not accompanied in Mary's thoughts by any representation of the letter, then Mary cannot be said to have grasped the relevant proposition. If John himself doesn't have any representation of the object he's talking about, his utterance cannot be regarded as a speech act. That is why unarticulated constituents of the content are needed.

Since I – contrary to Recanati – propose unarticulated constituents in the explicit content, my account is a version of contextualism. Such a version might be called situated contextualism, because it stresses the importance of the situation that the utterance concerns: the elements of the situation are included both in the content and in the circumstances of evaluation. Recanati argues against unarticulated constituents in the content and, at the same time, against the Reflexivity Constraint, which says that the index with respect to which the relativized proposition is evaluated has to be the index of content (see Recanati 2007, p. 25). Thus, according to him, unarticulated constituents may be determined by speakers' intentions and need not be elements of the situation of utterance, but they all belong to the circumstances of evaluation. I think that in the case of subsentential speech acts the Reflexive Constraint holds, i.e. contents of subsentential utterances have to be evaluated relative to the index of the content, but some of the elements of the situation must be cognitively articulated in beliefs expressed by subsententials (and thus be things that the proposition expressed is about). If they were not cognitively articulated (or if – as Recanati argues – they were articulated within other beliefs held by the speaker), subsentential utterances couldn't be regarded as speech acts, because their content would be too impoverished. This is especially evident if what is uttered is the expression whose content is an argument for the propositional function that needs to be supplemented by the situation. For instance, if the speaker utters "Slab" but has no action involving slabs in mind, we cannot say that he made a speech act only on the basis of the fact that the context determines some such actions.

Recanati credits Perry with the idea of distribution, and I think it is Perry to whom we should turn for help with rebutting the objection from connectivity. In "Davidson's Sentences and Wittgenstein's Builders", Perry introduces a distinction between constructing a complete thought and completing a thought. One constructs a complete thought when one "utters a sentence that identifies all the necessary propositional constituents for a thought" (Perry 1994, p. 33). One completes a thought when one introduces "the last needed constituent to form a proposition" (Perry 1994, p. 33). Thus, a builder may either construct a complete thought by saying "Pass me a slab" or he may just say "Slab" and complete a thought whose other

constituents are already present in the situation.[13] In my view, one might argue that case matching is a tool that makes completing a thought easier for the hearer. When the speaker utters a phrase with the intention of introducing the constituent needed to complete a thought, usually there will be more than one salient way of interpreting his utterance by using this constituent to complete a thought.

Perry quotes Wittgenstein's remark 6 from *Philosophical Investigations*:

> "I set the brake up by connecting up rod and lever." - Yes, given the whole of the rest of the mechanism. Only in conjunction with that is it a brake-lever, and separated from its support it is not even a lever; it may be *anything*, or nothing.
>
> (Wittgenstein 1953, § 6, my italics).

In order to make the task of interpreting the utterance easier for the hearer, in order to make it evident how to connect the rod to the rest of the mechanism, in the appropriate circumstances the speaker may use an expression in such a grammatical case as to make it clear how his utterance is to be completed. 'Completing the thought' is, on the pragmatics-oriented view, a task involving pragmatic inferences, thus it involves some 'guess-timating' – reading the speaker's intentions and particular grammatical cases may help eliminate certain possibilities. When it is not particularly important how the utterance will be understood (e.g. when someone says "Ładna sukienka" (Nice dress. NOM) it is not essential to the speaker whether he will be interpreted as saying "To jest ładna sukienka" (This is a nice dress) or "Masz ładną sukienkę" (You're wearing a nice dress)), the speaker will choose the nominative case. However, when there might be some uncertainty regarding how to interpret the utterance (e.g. in the father and clumsy daughter case), the speaker chooses the case that makes the intended interpretation more determinate. One might wonder whether this view is able to explain why in some cases speakers cannot use grammatical cases other than the nominative case (e.g. why Hans cannot say "An meinen Vater" or a person seeing a friend wearing a new dress cannot utter "Ładną sukienkę"), even though the relevant utterances are fragments of sentences that could have been uttered. A possible – although probably not fully satisfactory – explanation is that inter-

---

13 In this respect Perry's view is similar to Stainton's, who argues that the hearer recognizes that the speaker intends to communicate a proposition but at the same time notices that what was uttered has non-propositional meaning. Hence, the hearer starts looking for the missing argument or function that would complete the proposition expressed according to the intentions of the speaker. See Stainton (2006a, p. 158).

preting such utterances by recovering the sentences that could have been uttered, is more costly and less efficient than supplementing them directly with elements from the context.[14]

6 Conclusion

The semantics-oriented view which regards alleged subsentential speech acts as ellipses faces a serious objection from indeterminacy. Unless this objection is disarmed, the ellipsis view cannot be regarded as a viable account for explaining subsentential speech acts away. On the other hand, the most powerful objection against the pragmatics-oriented view is the argument from connectivity. I have endeavoured to argue that one may resist the force of this argument if one adopts a particular version of the view, namely the account that I call situated contextualism, which is based on Recanati's moderate relativism. On this view, the speaker relies on the discourse situation to cooperate and supply the missing ingredients of the thought he is trying to express. Since usually there is more than

---

14 The anonymous reviewer has pointed out that my solution to the problem of explaining connectivity effects is similar to that proposed by Begoña Vicente and Marjolein Groefsema in "Something out of nothing? Rethinking unarticulated constituents" (*Journal of Pragmatics* 47 (2013), pp. 108–127). Indeed their solution (of which I was unaware at the time of writing the paper) is similar in spirit, but it is a part of an importantly different view on how subsententials are interpreted. First of all, Vicente and Groefsema argue that "the notion of unarticulated constituent (...) should be abandoned" (2013, p. 111). They claim that "the linguistic input maps directly onto an evolving logical conceptual representation whose well-formedness conditions it has immediate access to. This means, crucially, that there is no need to postulate a level of representation of linguistically articulated content other than the conceptually articulated content that the language processor builds online". (2013, p. 110). In other words, all allegedly unarticulated constituents are in fact conceptually articulated ones. Moreover, while on my account – as on Stainton's – the hearer arrives at the proposition expressed by combining the content from language with the content from the discourse situation, on Vicente and Groefsema's view the linguistic input is directly mapped to the representation in the language of thought. Composition rules operate directly over language of thought representations and not over linguistic representations (2013, p. 120). Their explanation of case matching is that cases "introduce constraints on the structural position of the content of the NP, which must be respected whether or not the whole of the input is provided linguistically" (2013, p. 120). Thus, the underlying language of thought proposition fragment activated in the situation determines the appropriate case of the fragment (ibid.). The speaker doesn't have a determinate content in mind, but, by using the case, restricts the range of possible completions for the hearer.

one salient way of completing the thought, the speaker may revert to using grammatical cases other than the nominal case in order to simplify the hearer's task. Thus, I think the situated contextualist account is able to withstand the objection from connectivity and can be regarded as a view that explains how it is that subsentential expressions can be used to make a move in a language game.[15]

References

Bach, Kent (2008): "Review: Robert J. Stainton: *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language.*" In: *Mind* 117, p. 739–742.

Barton, Ellen (1990): *Nonsentential Constituents: A Theory of Grammatical Structure and Pragmatic Interpretation*, Amsterdam: John Benjamins.

Corazza, Eros (2011): "Unenriched Subsentential Illocutions." In: *Philosophy and Phenomenological Research* 83, p. 560–580.

Elugardo, Robert/Stainton, Robert (2004): "Shorthand, Syntactic Ellipsis, and the Pragmatic Determinants of What Is Said." In: *Mind and Language* 19 (4), pp. 442–471.

Elugardo, Robert/Stainton, Robert (2005): "Introduction." In: Robert Elugardo/Robert Stainton (Eds.), *Ellipsis and Nonsentential Speech*, Dordrecht: Springer, pp. 1–26.

Hall, Allison (2009): "Subsentential Utterances, Ellipsis, and Pragmatic Enrichment." In: *Pragmatics and Cognition* 17(2), pp. 222–250.

Korta, Kepa/Perry, John (2011): *Critical Pragmatics. An Inquiry into Reference and Communication*. Cambridge: Cambridge University Press.

Martí, Luise (2015): "Grammar versus Pragmatics: Carving Nature at the Joints." In: *Mind and Language* 30, pp. 437–473.

Merchant, Jason (2010): "Three Kinds of Ellipsis: Syntactic, Semantic, Pragmatic?" In: François Recanati/Isidora Stojanovic/Neftali Villanueva (Eds.): *Context-Dependence, Perspective, and Relativity*. Berlin: The Gruyter Mouton, pp. 141–192.

Merchant, Jason (2004): "Fragments and Ellipsis". In: *Linguistics and Philosophy* 27, pp. 661–738.

Merchant, Jason (2006): "Small Structures: A Sententialist Perspective". In: Ljiljana Progovac/Kate Paesani/Eugenia Casielles-Suarez/Ellen Barton (Eds.): *The Syntax of Nonsententials: Multidisciplinary Perspectives*. Amsterdam/Philadelphia: John Benjamins, pp. 73–91.

Odrowąż-Sypniewska, Joanna (2017): "Are There Any Subsentential Speech Acts?" In: *Studia Semiotyczne – English Supplement* 29, pp. 248–271.

Odrowąż-Sypniewska, Joanna (2018): "Subsentential Speech Acts: a Situated Contextualist Account". In: Tadeusz Ciecierski/Paweł Grabarczyk (Eds.): *The Architecture of Context and Context-Sensitivity. Studies in Linguistics and Philosophy,* Springer (forthcoming).

Perry, John (1986): "Thought Without Representation." In: *Proceedings of the Aristotelian Society, Supplementary Volumes* 60, pp. 137–151.

Perry, John (1994): "Davidson's Sentences and Wittgenstein's Builders." In: *Proceedings and Addresses of APA* 68, pp. 23–37.

Progovac, Liljana (2006): "The Syntax of Nonsententials: Small Clauses and Phrases at the Root". In: Ljiljana Progovac/Kate Paesani/Eugenia Casielles-Suarez/Ellen Barton (Eds.) *The Syntax of Nonsententials. Multidisciplinary Perspectives*. Amsterdam: John Benjamins Publishing Company, pp. 33–72.

Recanati, François (2007): *Perspectival Thought. A Plea for (Moderate) Relativism*. Oxford: Oxford University Press.

Stainton, Robert (2006a): *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford: Oxford University Press.

Stainton, Robert (2006b): "Neither Fragments Nor Ellipsis." In: Ljiljana Progovac/Kate Paesani/Eugenia Casielles-Suarez/Ellen Barton (Eds.): *The Syntax of Nonsententials: Multidisciplinary Perspectives*. Amsterdam/Philadelphia: John Benjamins, pp. 93–115.

Stanley, Jason (2000): "Context and Logical Form." In: *Linguistics and Philosophy* 23, pp. 391–434.

Wittgenstein, Ludwig (1953): *Philosophical Investigations*. Translated by Elizabeth Anscombe. Oxford: Blackwell.

Maciej Witek

Accommodation in Linguistic Interaction. On the So-called Triggering Problem

Abstract: Accommodation is a process whereby the context of an utterance is adjusted or repaired in order to maintain the default assumption that the utterance constitutes an appropriate conversational move of a certain type. It involves a kind of redressive action on the part of the audience and, depending on what the appropriateness of a speech act requires, results in providing missing contextual elements such as referents for anaphoric expressions, presuppositions, suppositions, deontic facts, pragmatically enriched contents, and so on. It remains to be determined, however, what is the source of the contextual requirements whose recognition motivates and guides the accommodating context-change. The aim of this paper is to address this question – which expresses the so-called triggering or constitution problem – and suggest that it can be adequately answered by a speech act-based model, whose central idea is that the requirements in question are structural components of patterns, scripts or procedures for the performance of speech acts.

The paper consists of four parts. Section 1 introduces the notion of accommodation and discusses three examples of accommodating phenomena. Section 2 develops a more elaborated description of the examples discussed in the previous section and proposes a list of questions that an adequate model of accommodation is expected to answer. Section 3 offers a critical examination of three alternative models of accommodation, i.e., David Lewis's score-keeping model, Robert Stalnaker's sequential update model, and Richmond Thomason's enlightened update model; in particular, it considers how they account for the constitution of contextual requirements that trigger and guide mechanisms of context-redressive changes. Finally, Section 4, suggests basic elements of a speech act-based model; it also argues that the proposed framework can be used to explain a wide range of accommodating phenomena and can shed a new sort of light on the constitution of accommodation-triggering requirements. [1]

1 Introduction

Accommodation is a cooperative response to the requirements that the performance of a speech act imposes on the context relative to which the act is to be evaluated and interpreted; to put it differently, it is a process whereby the context of an utterance is adjusted or repaired so as to maintain the default assumption that the utterance constitutes an appropriate conversational move. Consider, for example, the following utterance:

(1) a. Jacqueline's getting married.
     b. He's a soccer coach.[2]

The use of the anaphoric pronoun 'he' requires, for its appropriateness, that in the set of available referents there is a contextually salient male person to whom it refers. Knowing that 'Jacqueline' is a female name, however, we cannot regard it as a discursive antecedent providing information relevant for interpreting the token of 'he' under discussion. What is more, the discursive context of (1b) involves no explicit mention of a male person that could be regarded as a referent for this anaphoric expression. Nevertheless, the gap can easily be filled and the missing element can readily be provided; in other words, the context can be adjusted through accommodation so as to meet the requirement imposed by the anaphoric use of 'he'. Taking into account the meaning of sentence (1a), namely, it is natural to assume that in uttering (1b) the speaker expresses the proposition that *Jacqueline's fiancé* is a soccer coach.[3]

---

2 I borrow this example from Roberts (2015).

3 By saying that it is natural to assume this interpretation of dialogue (1), I do not mean that there are no other readings of this interaction — as well as of other interactions examined in this paper — available to different audiences. I am grateful to an anonymous reviewer for drawing my attention to this point.

Consider, by analogy, a situation in which a few passengers of a cruise have survived a marine disaster and found themselves on a desert island. Everybody is shocked. Only Jones keeps a cool head and, taking the initiative, says to Smith:

(2) Go and pick up wood![4]

Let us assume that Smith complies with what he is told and that next utterances of imperative sentences that Jones addresses to the survivors are taken to be binding orders. Note, however, that the appropriateness or felicity of an order *qua* an illocutionary act requires that the speaker stands in a certain authority relation to his or her audience. In particular, the felicity of the order made in the utterance of (2) requires that Jones has authority over Smith. Let us assume, however, that this requirement was not fulfilled prior to Jones's utterance. Does it make his order an Austinian misfire, i.e., an act purported but "null and void" (Austin 1962, p. 25)? Not necessarily. The context can be repaired by accommodation. More specifically, Jones's authority can be produced indirectly through a context-fixing mechanism guided by the assumption that his utterance of sentence (2) is *taken to be* and *takes effect as* a binding order.

Finally, consider a situation in which Phoebe utters the following two sentences:

(3) a. I cannot come to the meeting.
     b. I have to pick up my cat at the veterinarian.[5]

Let us assume that prior to the utterance Phoebe's hearers had no idea whether or not she had a cat. Nevertheless, after hearing her words they are entitled to accept the proposition that Phoebe has a cat. The proposition, however, is not asserted but presupposed; more specifically, it is triggered by the possessive nominal phrase 'my cat' occurring in (3b), whose appropriate use seems to require that the proposition is part of the common ground relative to which the assertion made in uttering (3b) is to be interpreted. To meet this requirement, the common ground between Phoebe and her interlocutors – i.e., their "mutually recognized shared information in a situation in which an act of trying to communicate takes place" (Stalnaker 2002, p. 704) – is automatically updated or adjusted so as to accommodate the proposition that Phoebe has a cat.

---

4 This is a variation on an example originally discussed by J.L. Austin; see Austin (1962, p. 28); cf. Langton (2015, pp. 2–5); Witek (2013, p. 154) and Witek (2015b, p. 14).
5 For a discussion of this example, see Stalnaker (1998, p. 9); von Fintel (2008, p. 144).

Accommodation, then, comes in different forms. It involves "some redressive action on the part of the addressee" (Simons 2003, p. 258) and, depending on what the appropriateness of a speech act requires, can provide missing contextual elements such as referents for anaphoric expressions, deontic powers, and presuppositions.[6] It is not clear, however, what is the source of the contextual requirements whose recognition motivates and triggers the accommodating context-change. My aim in this paper is to address this question – which expresses the so-called *triggering problem*[7] – and suggest that it can be adequately answered by a speech act-based model, a central tenet of which is that the requirements in question are structural components of patterns, scripts or procedures for the performance of speech acts.

The remaining part of this paper is organized into three parts. Section 2 offers a more detailed discussion of the above-mentioned examples of context-redressive mechanisms and proposes a list of questions that an adequate model of accommodation is expected to address. Next, Section 3 discusses three alternative models of accommodation – i.e., David Lewis's (1979) score-keeping model, Robert Stalnaker's (1998, 2002, and 2014) sequential update model, and Richmond Thomason's enlightened update model[8] – and considers how they account for the nature and dynamics of accommodation as well as the constitution of the contextual requirements that trigger and guide mechanisms of context-redressive changes. Finally, Section 4, which is the concluding part of the paper, proposes basic elements of a speech act-based framework within which one might explain a wide range of accommodation phenomena; in particular, the proposed framework can throw a new light on the constitution of accommodation-triggering requirements.

---

6 Laura Caponetto (in personal communication) and an anonymous reviewer have pointed out that elements produced or modified by context-redressive mechanisms — e.g. referents for anaphoric expressions, presuppositions, deontic powers, and so on — fail to form a homogenous class and, as the corollary of this, it is doubtful whether there is a universal accommodation process underlying all the phenomena under scrutiny. I do agree with this remark. I am far from saying that there is a common denominator underlying all the varieties of accommodating mechanisms considered in this paper. I claim, rather, that the speech act-based model offers a unified framework for *describing* the structure of accommodation phenomena; to *explain* concrete instances of accommodating mechanisms, however, one has to allow for the particularities of the language games within which they occur (for instance, for a discussion of accommodation of deontic powers in illocutionary games, see Witek 2013 and 2015b).

7 See Simons (2001, p. 431); von Fintel (2008, p. 138); Domaneschi (2017).

8 See Thomason (1990) and Thomason *et al.* (2006).

2 What we want to know about accommodation and are not afraid to ask

Following the speech-act theoretic tradition, let us assume that speech acts are "context-changing social actions" (Sbisà 2002, p. 421). For example, a binding promise changes a context in which the speaker is not committed to perform the act she predicates of herself into a context in which she is so committed; by analogy, a felicitous order changes a context in which the addressee is not obliged to perform a certain act into a context he is so obliged. According to the common-ground theory of assertion[9], in turn, a successful assertion updates the common ground with its propositional content, provided that the hearer accepts it, or at least acknowledges it with the proposition that the assertion has been made.

Viewed from this perspective, accommodation seems to involve a special kind of context-changing mechanism, which operates indirectly and implicitly rather than directly and explicitly[10]; in other words, it exploits what Rae Langton calls *back-door* methods[11] rather than direct or *official* mechanisms of context-change. For instance, the order made in uttering sentence (2) affects the context of its performance by adding to it the following three normative facts:

(4) a. Smith's *obligation* to pick up wood,
     b. Jones's *right* to expect Smith to pick wood,
     c. Jones's *authority* over Smith.

Note, however, that only facts (4a) and (4b) can be regarded as direct effects of Jones's order; by contrast, fact (4c) should be viewed as one of the deontic presuppositions of the illocutionary act under discussion. In other words, the felicity of Jones's order *qua* order *results* or *consists in* producing facts (4a) and (4b) and *presupposes* or *requires* fact (4c). If the requirement is not satisfied by prior context, it comes into being through a 'back-door' or indirect mechanism of accommodation.

Consider, next, the utterance of (1). Let us assume that it affects the context of its production – construed of as the common ground – by adding to it the following four propositions or contents:

---

9 See Stalnaker (2002); for a critical discussion, see Abbott (2008).

10 For a discussion of indirect mechanism of illocutionary interaction, see Witek (2015b).

11 See Langton *forthcoming b*.

(5) a. Jacqueline's getting married.
     b. [A certain contextually salient male person] is a soccer coach.
     c. Jacqueline is getting married to a certain male person, who is her fiancé.
     d. Jacqueline's fiancé is a soccer coach.

Proposition (5a) can be regarded as representing what the speaker of (1a) literally says. Content (5b), in turn, is the literal meaning of sentence (1b). It is not propositional, since it contains a procedural component "[A certain contextually salient male person]" which is the linguistically determined meaning of the pronoun 'he'; namely, according to the grammar of English 'he' can be used literally to refer to a certain contextually salient male person.[12] Proposition (5c) can be regarded as a bridging assumption whose function is to provide a discursive referent for the pronoun 'he' used in sentence (1b).[13] Proposition (5d), in turn, represents what the speaker says in uttering sentence (1b); more specifically, it results from interpreting the utterance of sentence (1b) – whose linguistically determined meaning is represented by (5b) – against the background of bridging assumption (5c). In sum, we can take contents (5a) and (5b) to result from literal interpretation of sentences (1a) and (1b), respectively; but the complete interpretation of the speech act made in uttering these sentences involves propositions (5c) and (5d) which are determined through accommodation: bridging assumption (5c) provides a discursive referent required by the use of pronoun 'he', thereby contributing to the determination of proposition (5d).

Let us also re-consider the situation in which Phoebe utters sentences (3a) and (3b), thereby asserting that she cannot come to the meeting and that she has to pick up her cat at the veterinarian. Provided that her interlocutors accept or ground her assertions, the common ground is updated with the following five propositions that are parts of the total content of Phoebe's utterance:

(6) a. Phoebe cannot come to the meeting.
     b. Phoebe has to pick up her cat at the veterinarian.
     c. Proposition (6b) explains proposition (6a).

---

12 More specifically, the condition in square brackets is a *referential constraint* in the sense introduced by Kent Bach; see Bach (1987, pp. 186–188; 2001, p. 33); cf. Witek (2015a, p. 19).
13 For a discussion of bridging assumptions and the role that they play in interpreting anaphoric expression see Wilson and Matsui (2012).

    d. The time of the meeting mentioned in (3a)
       is the time of the visit at the veterinarian mentioned in (3b).
    e. Phoebe has a cat.

Propositions (6a) and (6b) can be regarded as contents of two assertions that Phoebe makes in her utterances of sentences (3a) and (3b), respectively. Propositions (6c) and (6d), in turn, go beyond what she literally says and as such can be regarded as what she conversationally implicates. It is possible, however, to regard them as elements of linguistically determined meaning of the complex utterance (3). Following Asher and Lascarides (2003), one can assume that what makes (3) a coherent discourse is the fact that its constituent segments (3a) and (3b) stand in a certain rhetorical relation to each other, i.e., the rhetorical relation of *Explanation* reported in proposition (6c), and that their being so related evokes predictable "truth conditional effects on the content of the discourse" (Asher and Lascarides 2001, p. 202), e.g., the effect reported in proposition (6d). In short, propositions (6c) and (6c) can be regarded as determined by the semantics of rhetorical relations which constitutes an element of what can be called, following Asher and Lascarides, the *extended notion of grammar*[14]. Finally, proposition (6e) is what Phoebe *presupposes* or *supposes* in her utterance of sentence (3b). In fact, this presupposition – which is triggered by the possessive nominal phrase 'my cat' – is informative; according to some scholars – e.g., Robert Stalnaker, Kai von Fintel, and others[15] – the functioning of informative presuppositions is regarded to involve accommodation construed of as a mechanism whose job is to adjust the common ground so as to make the presupposition satisfied.

    Generally, accommodation can be described as a mechanism whereby the context of a speech act is redressed – repaired and fixed or merely adjusted – in order to maintain the default presumption that the act is an appropriate conversational move of a certain type: a felicitous illocution, an appropriate act of anaphoric reference, a satisfied presupposition, and so on. In other words, the appropriateness of a speech act puts certain requirements on the context in which it is performed, and the recognition of these requirements triggers and motivates a cooperative response of the audience that results in an implicit and indirect context-change. This general description is true as far as it goes. In what follows,

---

14 See Asher and Lascarides (2001, p. 210); cf. Chapter 6 of Lepore and Stone (2015), and a discussion of the idea of *extended semantics* in Witek (2016).

15 See Stalnaker (2002, p. 711), and von Fintel (2008, pp. 140–141); for a critical discussion of the common ground account of informative presuppositions see Gauker (1998, pp. 160–162), and Abbott (2008, pp. 529–531).

however, I would like to go further and consider the possibility of developing a comprehensive framework that might be used to allow for varieties of accommodating phenomena. Accommodation comes in many forms that can differ with respect to:

i. the nature of their underlying mechanisms (e.g., accommodation can be construed either as a rule-governed process, a special case of cooperative transaction, or abductive intention-recognition inferences),

ii. the kind of context that they affect (e.g., accommodated context can be represented either as a domain of normative facts, a universe of available discourse referents in the sense of the Discourse Representation Theory[16], or the common ground),

iii. the kind of redressive action they involve (e.g., accommodation can involve either context-repair or context-adjustment),

iv. the kind of inappropriateness that could ensue without it (e.g., accommodation failures can result in infelicitous illocutions, referential failures, unsatisfied presuppositions, and so on[17]),

v. the source of the requirements whose recognition initiates and motivates the indirect and implicit process of context-change.

Therefore, it is reasonable to expect any adequate model of accommodation to address and answer the following five questions:

Q_1 What is the nature of accommodating mechanisms?

Q_2 What type of context does accommodation affect?

Q_3 What kind of redressive action does it involve?

Q_4 What sort of inappropriateness would ensue without it?

Q_5 What is the nature and source of the contextual requirements whose recognition motivates the accommodating context-change?

My aim in the next section is to discuss three received models of accommodation – developed, respectively, by David Lewis, Robert Stalnaker and Richmond Thomason – and consider how they address the above-mentioned questions. Before getting into the details, however, it is worth noting that question Q_5 is to be read

---

16 See Guerts *et al.* (2016).

17 Of course this list of failures is far from being complete and requires further elaboration. For example – as Laura Caponetto has pointed out to me (in personal communication) – some illocutionary misfires result from referential failures or other semantic defects, whereas other misfires occur as a consequence of the speaker's lack of a required authority.

as concerning the constitution of contextual requirements rather than their recognition, retrieval and processing; in other words, the question is *metaphysical* or *constitutive* rather than *epistemic*. The contrast between the constitutive and epistemic readings of a theoretical claim or question comes from Bach, who uses it to distinguish between two senses of the verb 'determine' occurring in statements of the form '*X* determines the content of utterance *U*'. On their epistemic reading, those determination-statements are to be paraphrased as '*X* ascertains the content of utterance *U*'; on their constitutive reading, by contrast, they are to be paraphrased as '*X* constitutes the content of utterance *U*'.

> Determination of the first sort is *epistemic* determination. It does not make it the case that an utterance has a certain content or, in particular, that the speaker has said a certain thing. Rather, it is what is involved in the hearer's figuring these things out, as well as what the speaker is communicating. *Constitutive* determination makes it the case that the speaker has said a certain thing and, more generally, that an utterance has a certain content.
>
> (Bach 2001, p. 29–30)

In what follows I assume that the requirements that the performance of a speech act puts on its context are part of the act's content. I also take question Q_5 to be constitutive rather than epistemic; I assume, namely, that to answer it – and thereby to solve the triggering problem – is to identify those aspects of (the structure of) a speech situation under scrutiny that *constitute* the requirements the recognition of which motivates the accommodating context-change. As Langton (2015, p. 1) puts it, in accommodation "what is said 'requires and thereby creates' what is required"; it remains to be determined, however, what is the source of the requirements in question: how they arise or, in other words, how they are constituted.

3 Three models of accommodation

3.1 Lewis's score-keeping model

A central idea behind Lewis's model of conversational dynamics – or, more accurately, of conversational kinematics[18] – is that *context* can be represented as *conversational score*.[19] The conversational score at a given stage of a language game is a sequence of abstract entities that represent those aspects of the context of a speech act relative to which the act is to be interpreted and evaluated; it can also be defined as an abstract data structure whose function is to track and represent publicly recognizable contributions to the state of the conversation. These two complementary definitions capture two roles that the notion of conversational score is designed to play. First, it plays an interpretative or evaluative role in that score constitutes the background relative to which conversational moves are to be interpreted or evaluated; in other words, the conversational score at a given stage of a game constraints the scope of moves that can be appropriately made at this stage. Second, it plays a kinematic role in that it evolves in response to the moves made in the game.

According to Lewis, then, the score of a language game involves components that, first, are relevant for interpreting its constituent moves and, second, change in response to what happens during the game. They can be presuppositions shared by interlocutors, deontic fact or normative relations (e.g. commitments, obligations, rights, entitlements, and so on), discourse referents, rankings of comparative salience, standards of precision, and so on.

Lewis's key idea is that the kinematics of conversational score is a rule-governed process. He distinguishes between two types of what he calls 'rules of score-change' or 'kinematics rules': *rules of direct kinematics* and *rules of accommodation*. The former are constitutive in that they determine how moves of a certain type affect the context of their performance; in other words, they specify, for any move that can be appropriately made in the game, what would count as its target score. Therefore, move types – as well as the rules of direct kinematics that

---

18 In personal communication Marina Sbisà has noted that conversational dynamics concerns utterances construed of as speech acts or actions, whereas conversational kinematics concerns utterances understood as moves made in a language game. Undoubtedly Lewis was interested in conversational kinematics rather than dynamics.

19 Lewis (1979); c.f. Kölbel (2011), and Langton forthcoming a, b; what Lewis calls *conversational score* can be to some extent likened to *conversational record* is the sense introduced by Thomason (1990) and further elaborated by Lepore and Stone (2015).

are constitutive of them – can be represented as functions from source-scores into target-scores. Rules of accommodation, by contrast, govern a process whereby the context of a move is adjusted so as to make the move appropriate; even though they can be represented as functions from scores into scores – more specifically, as functions from *defected* source-scores to *accommodated* or *repaired* source-scores – they do not play a constitutive role.

Lewis proposes the following general pattern for rules of accommodation:

> If at time t something is said that requires component s_n of conversational score to have a value in the range *r* if what is said is to be true, or otherwise acceptable; and if s_n does not have a value in the range *r* just before *t*; and if such-and-such further conditions hold; then at *t* the score-component s_n takes some value in the range *r*.
>
> (Lewis 1979, p. 347)

Accommodation, then, can operate only if certain conditions are met, e.g., if nobody blocks it.[20] It is also instructive to note that Lewis takes conversational moves to be acts of saying and takes the truth of what is said to be a designated aspect of its appropriateness. Of course Lewis allows for other forms of appropriateness. According to him, however,

> [w]e need not ask just what sort of unacceptability results when a required presupposition is lacking. Some say falsehood, some say lack of truth value, some just say that it's the kind of unacceptability that results when a required presupposition is lacking; and some say it might vary from case to case.
>
> (Lewis 1979, p. 739)

Nevertheless, Lewis takes it that in most cases what motivates the adapting context-change is the need to maintain the default assumption that what the speaker says is true. Consider, for example, a master-slave game whose constituent moves can be described as Austinian exercitives, i.e. acts whose function is to enact new rules, norms and permissibility facts; exercitives presuppose also certain felicity conditions: they are "the exercising of powers, rights, or influence" (Austin 1962, p. 150). Let us imagine a master who says to a slave:

(7) You are now permitted to cross the white line.[21]

---

20 For a discussion of blocking see Langton *forthcoming b*.

21 For a more extensive discussion of this example, see Langton *forthcoming a*.

thereby exercising his power and creating a new permissibility fact or norm. According to Lewis, the mechanism underlying the production of this new norm involves accommodation. In other words, permissibility facts follow a rule of accommodation:

> If at time t something is said about permissibility by the master to the slave that requires for its truth the permissibility or impermissibility of certain courses of action, and if just before t the boundary is such as to make the master's statement false, then – *ceteris paribus* and within certain limits – the boundary shifts at t so as to make the master's statement true.
>
> (Lewis 1979, p. 341)

According to this rule, the context-adjusting process that results in creating new norms is motivated by the need to maintain the default assumption that whatever the master says is true.

Let us also consider a ceremonial act of naming a ship made with the help of the following explicit performative formula:

(8) I name this ship the Queen Elizabeth.[22]

The utterance of (8) affects the social context by creating a new conventional fact, i.e., that the indicated ship bears the name the *Queen Elizabeth*. According to Lewis, the production of this effect involves a process governed by the following rule of accommodation:

> If at time *t* something is said that requires for its truth that *X* bear name *n*; and if *X* does not bear *n* just before *t*; and if the form and circumstances of what is said satisfy certain conditions of felicity; then *X* begins at *t* to bear *n*.
>
> (Lewis 1979, p. 356)

According to the above-mentioned rule, then, the conversational move made in uttering sentence (8) should be regarded as an act of saying; what motivates the accommodating change is the assumption that what the speaker says is true.

In short, Lewis's account of exercitives and performatives is in a sense reductionist: he takes moves made in a language game to be acts of saying whose appropriateness consists in their being true. Consistently, Lewis accounts for the non-assertoric forces of utterances (7) and (8) in terms of indirect mechanisms of accommodation. In my view, in doing this he ignores the distinction between *openly made illocutions*, i.e. acts whose types are constituted by the rules of direct kinematics and represented as functions from source-scores into target-scores,

---

22 See Austin (1962, p. 116); cf. Langton *forthcoming a*.

and what Langton calls *back-door speech acts*, i.e., accommodation-exploiting acts that can be represented as functions from *defective* source-scores to *fixed* source-scores.[23]

Finally, let us consider Lewis's rule of accommodation for presuppositions:

> If at time *t* something is said that requires presupposition *P* to be acceptable, and if *P* is not presupposed just before *t*, then – *ceteris paribus* and within certain limits – presupposition *P* comes into existence at *t*.
>
> (Lewis 1979, p. 340)

If we assume that Phoebe's act of saying made in uttering sentence (3b) requires presupposition (6e) to be acceptable – i.e., to be part of the presuppositional component of the score – and that nobody blocks this presuppositions, it automatically becomes element of the repaired score. By analogy, let us assume that the utterance of (9) requires, for its appropriateness, presupposition (10):

(9) *Even* George could win.
(10) George is not a leading candidate.

If nobody objects by saying, for example 'Whaddya mean, *even*?'[24], then presupposition enters the conversation score. It is worth noting, however, that the 'back-door methods' exploited by accommodation have their limits. That is to say, even unblocked or unchallenged accommodation can fail to provide a required score component. Consider, for instance, the following utterance of sentence (11):

(11) Tonight, John is having dinner in New York, *too*.[25]

which contains the anaphoric presupposition trigger 'too'. Its appropriate use requires that the universe of available discourse referents contains a specific and contextually salient person (or group) who is having dinner in New York. If the discursive context of (11) involves no explicit mention of such a person, this gap can be filled with the help of an appropriate bridging assumption. However, if sentence (11) is uttered out of the blue, the available contextual information does not suffice to compute a required bridging assumption. Unlike the utterance (1)

---

23 For a criticism of Lewis's reductionism see Witek (2015b); for a defence and elaboration of the Lewisian model of illocutionary kinematics see Langton (2015).
24 See Langton *forthcoming b*.
25 See Kripke (1990).

discussed in detail in Section 2, the utterance of (11) fails to create or provide what it requires for its success.

Lewis's score-keeping model suggests at least partial answers to some of the questions formulated in Section 2. That is to say, he claims that accommodation is a rule governed process or, more specifically, a mechanism guided or even mandated by rules of accommodation (question Q_1); he assumes that accommodation affects conversation score (question Q_2) and involves context-repair or context-fixing construed of as providing missing contextual components (question Q_3); he also takes the falsehood of what the speaker says as a designated form of inappropriateness that would result without accommodation (question Q_4).

It is worth noting that Lewis fails to address question Q_5. In other words, he says nothing about the source of the contextual requirements whose recognition initiates accommodation. No wonder, since he is interested in the *kinematics* of presuppositions and other score components rather than in their *dynamics*; that is to say, his main concern is to track trajectories of context-changes rather than to account for them in terms of their underlying mechanisms and forces.

3.2 Stalnaker's model of sequential update

Unlike Lewis, Stalnaker limits his analyses to presuppositional requirements and presupposition accommodation. What is more, he represents context not as conversational score, but as common ground: a social object "definable in terms of the propositional attitudes of the members of some group" (Stalnaker 2014, p. 25). Nevertheless, the notion of common ground plays the role analogous to that of conversational score: first, it provides a background relative to which every new speech act is to be *interpreted*; second, it evolves in response to the speech acts made by the conversing agents as well as other manifest events they are mutually aware of. According to Stalnaker,

> [t]he notion of common ground in a propositional attitude concept. (...) it is a concept with an iterative structure: a proposition is common ground between you and me if we both accept it (for the purposes of the conversation), we both accept that we we both accept it, we both accept that we both accept that we both accept it, and so on.

> We can (...) define the individual propositional attitude of speaker presupposition in terms of common ground: An agent A presupposes that φ if and only if A accepts (for purposes of the conversation) that it is common ground that φ.

> (Stalnaker 2014, p. 25)

A key difference between Stalnaker's common-ground account and Lewis's score-keeping model is that Stalnaker claims that accommodation involves no rules of its own. He argues, namely, that:

> [a]ccommodation is an essential feature of any communicative practice. If common ground is (at least close to) common belief, then it will adjust and change in the face of manifest events that take place, including events that are themselves speech acts. Accommodation is just an example of this kind of change.
>
> (Stalnaker 2014, p. 58)

> A manifest event is something that happens in the environment of the relevant parties that is obviously evident to all. A goat walks into the room, or all of the lights suddenly go out. In such a case, it immediately becomes common knowledge that the event has happened (...).
>
> (Stalnaker 2014, p. 47)

The idea of speech acts as manifest events, then, plays a central role in Stalnaker's model of accommodation. For the sake of illustration, let us re-consider a situation in which Phoebe utters sentences (3a) and (3b). When sentence (3b) is uttered, it becomes a manifest event that a certain speech act with a certain set of properties has be made. In other words, it becomes part of the common ground among Phoebe and her interlocutors that in uttering (3b) Phoebe performs a speech act with (*i*) such and such a meaning and (*ii*) such and such a force; and (*iii*) presupposes that she has a cat; in other words, it becomes common ground among the conversing agents that Phoebe presupposes that she has a cat. According to Stalnaker, the later observation provides a key premise of a valid inference in the logic of common belief, whose conclusion is that the proposition that it is common ground among the agents that Phoebe has a cat.

It is not my aim to analyse and discuss the validity of Stalnaker's inference from the *premise* that it is common ground among the conversing agents that Phoebe presupposes that she has a cat to the *conclusion* that it is common ground among them that Phoebe has a cat.[26] My main concern in this paper is with the constitution of contextual requirements in general and presuppositional requirements in particular. Let us consider, then, what makes properties (*i*), (*ii*), and (*iii*) manifest aspects of Phoebe's utterance of (3b).

Let us assume that properties (*i*) and (*ii*) – i.e., the meaning and relational force[27] of Phoebe's utterance – are mutually recognized in virtue of the fact that

---

26 See Stalnaker (2002, pp. 709–711); for a discussion, see Simons (2003, pp. 261–262).

27 The relational force of the utterance of (3b) with respect to the utterance of (3a) can be identified with the rhetorical relation that holds between them, and that rhetorical relations can be

the extended semantics[28] of the language Phoebe speaks is common ground. That is to say, properties (*i*) and (*ii*) are *constituted* by rules of the extended semantics and, because the rules are part of the common ground among Phoebe and her interlocutors, are manifest aspects of the speech event she has produced. It is natural to consider, however, what *constitutes* property (*iii*) as a manifest aspect of Phoebe's utterance.

One possible solution to the above-mentioned *triggering problem*[29] is that the presuppositional requirement is a semantically determined property of the possessive nominal phrase 'my cat'; more generally, according to the so-called semantic view, presuppositional requirements "are hardwired in the semantics of particular expressions"[30]. It is worth noting that elements of the semantic view can be found in Stalnaker's paper "Common ground" published in 2002:

> If it is mutually recognized that a certain utterance type is standardly used, in some conventional linguistic practice, only when some proposition is (or is not) common belief, it will be possible to exploit this recognition, sometimes to bring it about that something is (or is not) common belief, sometimes to create a divergence between a conventionalized common ground and what speaker and hearer take to be the beliefs that they actually hold in common. The phenomenon of presupposition accommodation, much discussed in the literature about presupposition, is like the phenomenon of conversational implicature in that it is an inevitable feature of any practice the point of which is to mean things.
>
> (Stalnaker 2002, p. 705)

In short, Stalnaker seems to posit the existence of conventional rules that define standard or appropriate uses of presupposition triggers in terms of contextual requirements; the set of such requirements, in turn, seem to make up what Stalnaker calls the *conventionalised common ground*, which in the case of presuppositional failures can depart from the actual common ground. What is more, the conventions of standard uses can be exploited so as to bring about communicative effects akin to conversational implicatures.

---

regarded as linguistically determined aspects of the logical form of discourse (see Section 2 above and Asher and Lascarides 2001 and 2003).

28 For a discussion of the idea of extended semantics or extended grammar – which involves, apart from the rules of lexical and compositional semantics, the rhetorical structure rules – see Asher and Lascardies (2001), Chapter 6 of Lepore and Stone (2015), and Witek (2016).

29 For a formulation of the triggering problem, see Simons (2001, p. 431).

30 Von Fintel (2008, p. 138); apart from von Fintel, the proponents of the semantic view are Heim and Kratzer (1998) and Domaneschi (2017). For a similar account, see Witek (2016), where I distinguish between rules of appropriateness, construed of as part of the extended semantics, and the Maxim of Appropriateness, understood as a pragmatic norm.

Nevertheless, in his book *Context* published in 2014 he claims that "(...) presupposition requirements may have diverse explanations" (Stalnaker 2014, p. 70). Commenting on examples involving the use of anaphoric presupposition triggers, he claims:

> The simplest and most common case where a presupposition is required by the use of a sentence is a case where the addressee can apply the semantic rules to figure out what the speaker is saying only if he has certain information. In a case like this, the semantic rules help to explain why a presupposition is required, but the rules themselves need make no mention of presuppositions.
>
> (Stalnaker 2014, p. 53)

In other words, the determination of presuppositional requirements functions against the background of semantic rules and conventions, but the rules and conventions themselves make no reference to presuppositions. For the sake of illustration, let us re-consider the situation discussed in Section 3.1, in which the speaker out of the blue utters sentence (9). Recall that in this case the available contextual information does not suffice to compute a bridging assumption that could help determine a required discursive referent. What is, however, the source of this requirement? To answer this question, Stalnaker (2014, p. 71) refers to Heim's idea that 'too' means in addition to x'; in other words, the linguistically specified meaning of 'too' determines that its use involves tacit reference to a contextually salient x. Therefore, semantics helps to explain *why* a presupposition is required. In the case under discussion, it indicates what kind of contextual information we need to determine what the speaker says. Nevertheless, semantics as such fails to determine *what* exactly is presupposed by the use of 'too' in (9); in general, the determination or constitution of properties (*iii*) – i.e., of presuppositional requirements – is a pragmatic process.

According to Stalnaker, then, the performance of at least some speech acts construed of as context-changers gives rise to a sequential update that involves two steps: (*a*) accommodation understood as a cooperative response to a manifest speech event, i.e., to the mutual recognition that a certain speech act equipped with certain properties is made; and (*b*) the production of the 'essential effect' of the act, e.g., adding its content to the common ground. Step (*a*) leads us from *prior* common ground to *accommodated* common ground, whereas step (*b*) takes use from *accommodated* common ground to *updated* common ground. What mandates step (*a*) is a principle that Stalnaker (2014, p. 46) calls the *norm of agreement*, which is a variant of the Gricean Cooperative Principle.

The sequential update model is designed to solve the problem of informative presuppositions. However, it runs into a further trouble, i.e., it gives rise to the

so-called 'timing problem', whose discussion goes beyond the scope of the present paper[31]. Recall that my main concern is with the five questions formulated in Section 2. Let us consider, then, how they can be addressed and answered from the viewpoint of Stalnaker's sequential update model.

According to Stalnaker, accommodation involves a process guided by general principles of cooperation (question Q_1) and affects context construed of as common ground (question Q_2); it involves context-adjustment rather than context-fixing[32] (question Q_3). Finally, Stalnaker (2014, p. 63) claims that inappropriateness resulting from accommodation failures comes in many different forms (question Q_4); he maintains, however, that this fact poses no serious theoretical problem, since *appropriateness* is a descriptive rather than explanatory category, i.e. we use it merely to describe surface phenomena requiring explanation.

Like Lewis, Stalnaker is interested in how presuppositions behave and how they are accommodated rather than in how they arise. That is to say, apart from a few rather vague remarks on the source of some presuppositional requirements – e.g., those associated with the use of anaphoric triggers like 'too' occurring in sentence (9) – he gives no definite answer to question Q_5.

3.3 Thomason's model of enlightened update

Underlying Thomason's model of accommodation is the idea that interpretation is a kind of abductive intention recognition. He takes the intention behind a given act to be a complex information structure that involves (*a*) a goal, i.e., a state of affairs to be achieved, (*b*) a plan, i.e., a partially specified way of achieving the goal, and (*c*) preconditions, i.e., ways that the world is assumed to be, on which the achievement of the goal according to the plan depends.

Taking into account the structure of action-underlying intentions, we can describe an acting agent, first, as performing a certain public action, which is individuated by reference to the goal mentioned in (*a*), and, second, as making a series of tacit though publicly recognizable actions, which can be viewed as his or her undertaking a commitment or making a supposition that the preconditions mentioned in (*c*) are met. For example, a speaker who utters sentence (12):

---

31 For a discussion of the timing problem, see von Fintel (2008), Simons (2003) and Abbott (2008).
32 See Stalnaker (2002, p. 711); I put a discussion of the repair/adjustment contrast off until Section 3.3.

(12) Susan regrets that she bought a ferret.[33]

can be described as performing a *public action* of asserting that Susan regrets that she bought a ferret and a *tacit* though *publicly recognizable action* of supposing or committing herself to the claim that Susan bought a ferret.

Viewed from the perspective of the above-mentioned model of interpretation as intention recognition, accommodation is "a special case of obstacle elimination" (Thomason 1990, p. 343) and involves a kind of "enlightened update" (cf. Thomason et al. 2006). For the sake of illustration, let us re-consider the utterance of sentences (1a) and (1b). The hearer recognizes that the goal behind the speaker's utterance of (1b) is to assert that a certain male person is a soccer coach or, more precisely, to refer to a certain contextually salient male person and predicate the property of *being a soccer coach* of him. The hearer also recognizes that one of the preconditions for the achievement of this goal is that the universe of available discourse referents contains such a contextually salient object; however, he is also aware of the fact that the discursive context involves no explicit mention on a male person who could function as the referent for the anaphoric pronoun 'he' occurring in (1b). Being a cooperative interlocutor, however, he adopts the goal of eliminating this recognized obstacle and, consequently, accommodates the speaker by computing an appropriate bridging assumption that provides the missing referent and thereby repairs the context. Consider, by analogy, a situation in which prior to the utterance of (12) the hearer had no idea that Susan bought a ferret. He recognizes, however, that one of the preconditions for the achievement of the speaker's goal – namely, her asserting that Susan regrets that she bought a ferret – is that the proposition expressed by (13):

(13) Susan bought a ferret.

is part of the common ground of the conversation. Being a cooperative interlocutor, he adopts the goal of making this precondition satisfied and, consequently, accommodates the speaker by accepting her supposition that the proposition expressed by (13) is part of the common ground of the conversation. In short, to accommodate the speaker is to make the recognized preconditions for the achievements of her identified goals satisfied.

It is wort noting that Thomason's model allows us to view utterances (1) and (12) as initiating two different types of context-redressive actions. The mechanism

---

[33] I borrow this example from Craig Roberts; see Roberts (2015).

that provides the missing discursive referent for 'he' in (1b) can be best understood as a case of context-repair; in other words, it involves accommodation of presuppositions understood as requirements imposed on prior context. By contrast, the process that results in updating the common ground with the proposition expressed by (13) is to be understood as a case of context-adjustment; that is to say, it involves accommodation of suppositions rather than presuppositions. As Craig Roberts observes, *projective contents*[34] triggered by factive verbs and possessive nominal phrases – e.g., 'regret' in (12) and 'my cat' in (3b), respectively – are suppositions rather presuppositions; in other words, the appropriate use of the above-mentioned triggers does not require that the prior common ground entails the propositions expressed, respectively, by (13) and (6e), but that these propositions are not at issue relative to the current question under discussion[35].

In short, sentence (1b) is uttered in a context that fails to satisfy the requirements created put by the use of 'he' and, in this connection, is defective; that is why it needs repairing or fixing. By contrast, even though prior to her utterance of sentence (3b) Phoebe's interlocutor had no idea that she had a cat, her use of 'my cat' is appropriate as long as the proposition expressed by (6e) is not at issue relative to the current question under discussion. The same holds for the above-discussed utterance of (12): provided that the projective content expressed by (13) is not at issue relative to the current question under discussion, the use of 'regret' in (12) is appropriate no matter whether the hearer did or did not know that Susan had bought a ferret. Examples (3) and (12), then, involve accommodation understood as context-adjustment rather than context-repair: that is why the speakers' suppositions are accepted by the hearers "quietly and without fuss" (von Fintel 2008, p. 141).

Thomason, Stone and DeVault note that the above-presented account does not give rise to the problem of informative presuppositions. Therefore, we can give up the Stalnakerian idea of *sequential update* and replace it with the notion of *enlightened update*. Roughly speaking, it applies to cases "where the conversation moves forward not just through the positive effects of interlocutors' utterances but also from the retrospective insight interlocutors gain about one anothers' mental states from observing what they do" (Thomason et al. 2006, p. 5). In other words, enlightened update involves recognition and adoption of the

---

34 What makes these contents *projective* is that they persist in a number of embedding contexts, e.g., under negation, interrogation, a modal auxiliary, and so on. For a discussion, see Roberts (2015).

35 Roberts take questions under discussion to constitute one of the evolving components of conversational score; see Roberts (2015).

speaker's tacit though publicly recognizable commitments. According to Thomason, Stone and DeVault,

> [i]nformative presupposition arises as a problem in the presence of a pragmatic rule requiring an utterance involving a presupposition to be appropriate only if its presuppositions are mutually supposed at that stage of the conversation. We are not committed to such a rule; the alternative rules (...) would rather be (1) that an utterance involves a presupposition *P* if the intention underlying the utterance is committed to the presupposition, and (2) that an utterance is only appropriate to the extent that its presuppositions can be recognized and added to the common ground.
>
> (Thomason et al. 2006, p. 33)

Thomason also claims that accommodation is responsible for at least some of the interpretive effects that are traditionally described as conversational implicatures. Let us consider, for instance, Grice's *garage* example. Imagine a situation in which *A* is standing by an obviously immobilized car and is approached by *B*; the following talk-exchange takes place:

(14) *A*: I'm out of petrol.
      B: There is a garage around the corner.[36]

According to a standard Gricean reading of this example, in saying (*i*) that there is a garage around the corner *B* implicates (*ii*) that the indicated garage is open and has petrol to sell; in other words, content (*ii*) is worked-out through a pragmatic, maxim-guided inference. Viewed from the perspective of the enlightened update model, however, content (*ii*) can be regarded as determined by accommodation; more specifically, *A* recognizes that that the goal behind *B*'s utterance is to help him to find fuel, and that one of the preconditions for the achievement of this goal is that content (*ii*) is true or at least that *B* undertakes a tacit though publicly recognizable commitment to its truth.

Following Roberts, then, we can distinguish between three types of accommodation.

First, there are cases that, like (3b) and (12), involve context-adjustment rather than context-repair. They are associated with the use of projective content triggers and explicit expression of what is to be accommodated; the projective content is not presupposed, but supposed or taken for granted without being asserted; the accommodating mechanisms are linguistically mandated (i.e., trig-

---

36 Grice (1989, p. 32); cf. Thomason (1990, p. 347).

gered by the use of certain expressions or constructions) and linguistically controlled (i.e., the projective content is explicitly represented in the uttered lexical material).

Second, there are cases that, like (1) and (11), involve context-repair or context-fixing. They are associated with the use of presuppositional triggers that put conventional constraints on the kind of context in which they can be felicitously used; the accommodating mechanisms involved here are linguistically mandated, but are not linguistically controlled.

Third, there are cases that, like (14), involve the enrichment of the linguistically encoded meaning of an utterance. They are not associated with the use of overt triggers; their interpretation involves recognising and adopting preconditions abductively inferred to make sense of why and how the speaker is saying what she is saying; the accommodating mechanisms underlying such cases of *free enrichment* are neither linguistically mandated nor linguistically controlled.

Finally, let us consider how the enlightened update model addresses and elaborates on the triggering or constitution problem. It is instructive to note, first, that within the framework proposed by Thomason it takes the form of the following question: "What determines preconditions (i.e., private commitments and tacit actions) as parts of speakers' intentions?" One possible answer suggests that at least in some cases the preconditions can be determined by what Thomason calls the 'grammar':

> (...) the grammar might require a speaker to commit to certain information, privately, but in a publicly recognizably way, WITHOUT thereby requiring the speaker to somehow treat it as public, prior information. This gives an attractive way to resolve the well-known and frequent gaps where information must be grammatically backgrounded but need not be shared information among interlocutors. Classic examples include the informative presuppositions of change-of-state verbs, factives, and definite noun phrases.
>
> (Thomason *et al.* 2006, p. 6)

It is worth noting, however, that this *grammar-based* solution works only for cases of supposition accommodation that, like in examples (3b) and (12), involve linguistically mandated and linguistically controlled process. It remains to be considered, then, what determines the tacit actions involved in other types of accommodation. i.e., in accommodation of presuppositions and accommodation of enriching contents.

Let us conclude by gathering Thomason's answers to the five question formulated in Section 2. According to the enlightened update model, accommodation is cooperative process of obstacle elimination (question Q_1) and involves, depending on particular cases, either context-adjusting, context-fixing or content-enrichment (question Q_3); it can affect context construed of as conversational

score, whose components represent presuppositions, universes of discursive referents, and so on (question Q_2). The model under discussion is also pluralistic with respect to forms of inappropriateness that could result in cases of accommodation failures (question Q_4). Finally, it suggests that at least in some cases the contextual requirements whose recognition motivates accommodating context-changes are determined by the grammar (question Q_5).

4 Conclusions. Towards a speech act-based framework

Let us summarize the results of the previous sections. Recall, first, that we expect an adequate model of accommodation to answer a number of questions, some of which are listed in the concluding part of Section 2. They concern the nature and varieties of accommodating mechanisms (question Q_1), the sorts of contexts that they affect (question Q_2) and the kinds of redressive actions they involve (question Q_3), as well as the sorts of inappropriateness that would result from accommodation failures (question Q_4). In particular, we are interested in developing a comprehensive framework for explaining the constitution of the contextual requirements whose recognition motivates and triggers the accommodating context-changes (question Q_5); that is to say, we want to know what is the source of the requirements in question or, in other words, how they arise. This is the (in)famous triggering problem. The discussion presented in Section 3 suggests that the problem has not found a satisfactory solution yet. Lewis does not appear to address it at all. Stalnaker, even though in some parts of his works he seems to embrace the so-called semantic view – according to which contextual requirements on the appropriate use of prepositional triggers are semantically or conventionally determined – officially rejects it and claims that presupposition requirements have diverse, mostly pragmatic explanations. Thomason suggests that the requirements whose recognition initiates accommodation can be accounted for in terms of preconditions construed of as elements of action-underlying intentions or, more specifically, in terms of tacit though publicly recognizable actions or commitments that agents perform or undertake as part of their intentions to achieve certain goals in accordance with certain plans; but this suggestion, as it stands, needs further elaboration.

I am sympathetic to the idea that the contextual requirements whose recognition initiates accommodating mechanisms are built into the structure of our communicative plans and intentions. I also agree that at least in some cases –

especially the ones that involve the use of such triggers as factive verbs, posses-
sive noun phrases, and so on – the requirements in question are determined by
what Thomason calls the 'grammar'; more precisely, the not-at-issue projective
contents (6e) and (13) can be regarded as conventionally or semantically deter-
mined properties of sentences (3b) and (12), respectively. However, a similar ex-
planation applies neither to cases involving anaphoric triggers, whose pro-
cessing requires linguistically mandated though linguistically uncontrolled
processes, nor to cases involving no overt triggers, whose processing involves
free enrichment. What is more, it cannot be directly applied to the examples of
illocutionary acts made in uttering sentences (2) and (7), i.e., to accommodating
mechanisms that result in creating authority and permissibility facts, respec-
tively.[37]

My hypothesis is that the contextual requirements in question – which, de-
pending on particular cases, can be described as presuppositions, not-at-issue as-
sumptions, discursive referents, preconditions, tacit commitments, and so on – are
determined by rules of appropriateness for the performance of certain acts, i.e., by
rules that define the appropriateness of conversational moves made in a given lan-
guage game in terms of their *source scores*. In other words, following Lewis I as-
sume that the appropriateness rules determine, for any stage of the game, what
would count as a correct move at this stage. What is more, provided a given move
is taken by default to be appropriate, they help us determine and recognise its
contextual requirements and preconditions.[38]

It is instructive to note, following Sbisà (forthcoming), that in speech act the-
ory there are to alternative approaches to the study of appropriateness rules. For
the sake of the present paper let us call them the *Complete System View* and the
Incomplete System View. According to the former, the set of appropriateness rules
for the performance of acts of a given type is complete; therefore, it can be repre-
sented as part of a comprehensive list of necessary and collectively sufficient con-

---

37 What is more, it is doubtful whether the production of permissibility facts by exercitive illo-
cutionary acts involves accommodation at all; for a defence of such a Lewisian account of the
functioning of illocutionary acts, see Langton (2015); for a critical discussion, see (Witek 2015b).
38 It is instructive to stress that the proposed speech act-based approach does not presupposes
that there is a common denominator underlying all the varieties of accommodating phenomena.
What it offers is a general patterns for their description rather than explanation. As an anony-
mous reviewer has pointed out, it is doubtful whether accommodation can receive a clear gen-
eral account. In my view, accommodating phenomena comes in many different forms whose ad-
equate explanation requires allowing for the particularities of the language games within which
they occur. (see footnote 9 above and the discussion of the Incomplete System View below).

ditions for the performance of acts of this type. An example of the Complete System View is Searle's (1969) theory of illocutionary acts. According to the Incomplete System View, by contrast, the set of appropriateness rules for the performance of acts of a certain type is incomplete and open; therefore, even though we can explicitly formulated some of them, the resulting list is not closed. A classical formulation of the Incomplete System View can be found in Austin's lectures *How to Things with Words*[39]. As Sbisà notes, unlike Searle, Austin does not present his rules A, B, and Γ "as (templates for) jointly sufficient conditions, but leaves the performance of illocutionary act tokens open to unforeseen forms of defeasibility" (Sbisà *forthcoming*).

The speech act-based model proposed in this section is a variant of the Incomplete System View. I assume that studying breakdown cases we can discover new appropriateness conditions and thereby new requirements whose recognition can initiate accommodating mechanisms of context-change. As L.W. Forguson notes, the method of studying breakdown cases comes from Austin, who used it to justify his distinction between phatic, rhetic, locutionary, and illocutionary acts: "though they are merely abstractions from the fully happy, felicitous speech act, in 'breakdown' situations what is normally an abstraction may exist as an act in its own right to be ascribed to the speaker" (Forguson 1973, p. 165). In other words, "[a]lthough breakdowns may rarely occur in everyday speech situations, it is important to investigate them" (Forguson 1973, p. 171). In particular, studying them we can recognize new rules of appropriateness and unforeseen requirements that the performance of a speech act puts on its context.

A central idea behind the proposed speech act-based model is borrowed from Sbisà, who in her paper "Varieties of speech act norms" suggests that mechanisms underlying accommodation involve pattern-recognition and are guided by default assumptions of appropriateness. Like Stalnaker, Sbisà rejects Lewis's idea of accommodation as a rule-governed mechanism. Accommodation, she claims, involves no rules of its own. Unlike Stalnaker, however, she seems to take the idea of appropriateness to play a key role in determining the requirements the recognition of which can motivate context-redressive actions. In other words, accommodation functions against the background of speech act norms and rules, some of which are rules of appropriateness.

> Accommodation, then, is (...) governed by general principles, one of which concerns pattern recognition (a pattern can well be recognized from the presentation of some of its parts) and the other the by-default recognition of other minds or subjects. (...) It is indeed quite obvious

---

39 See Austin (1962); cf. Forguson (1973).

that a pattern that is partially presented may be completed by the observer if the part presented suffices to make it emerge.

(Sbisà forthcoming)

The existence of shared patterns of conventional action (which are cultural facts and can be expected to be linguistically encoded at least up to a certain point) could account for the "accepted conventional procedure" of Austin's rule A1 without binding us to an obsessionally rule-governed view of how illocutionary acts are performed. Moreover, patterns may be cognitively processed in different ways, for example by means of Gestalt-like mechanisms, but also, if needed (as in the case of unfamiliar patterns, gravely incomplete display, and other complications), inferentially, which would assign a legitimate role to inferential theories of illocutionary force understanding.

(Sbisà 2009, p. 48–49)

The above-mentioned ideas of pattern recognition and open systems of appropriateness rules can be used to fill the gap in the enlightened update model. Recall that according to Thomason action-underlying intentions are complex information structures that involves goals, plans and preconditions. My hypothesis is that the rules or appropriateness are built into the structure of such intentions and correlate goals and plans with the precondition of their achievement. Viewed from this perspective, accommodation involves, first, default evaluation of the speaker's utterance as an appropriate act of a certain type – i.e., as the act that achieves its goal construed of as either Austin's conventional effect or Searle's illocutionary point – and, second, the tacit assumption that all preconditions for the achievement of this goal are met (see Lewis's *ceteris paribus* condition).

How many rules of appropriateness are there? This is an empirical question that should be addressed and answered by studying breakdown cases. Taking into account the examples discussed in this paper, however, we can identify at least some of them. For example, there is a rule according to which the appropriate use of an anaphoric trigger requires that in the universe of available discourse referents there is a contextually salient individual of a certain sort (see examples (1) and (11)). Another rule says that the appropriate use of such projective content triggers as possessive nominal phrases and factive verbs require that the propositions they contribute are not-at-issue relative to the current question under discussion (see examples (3) and (12)). Consider, next, an act of providing a hearer with information that can help him achieve his signalled goals; one of the appropriateness conditions of this act is that the conveyed information is sufficiently rich (see example (14)). Finally, the felicity of an order requires that the speaker stands in an appropriate authority relation to her audience (see example (2)). One can also assume – in accordance with Lewis's reductionist model criticised in Section 3.1 above – that the appropriateness of explicit performative utterances,

as well as that of exercitives made in uttering indicative sentences about permissibility facts, consists in their being true (see examples (7) and (8)).

In sum, according to the proposed speech act-based model, accommodating mechanisms involve cognitive capacities akin to pattern-recognition (question Q_1) and can affect different components of conversational score (question Q_2); depending on particular cases, they can involve either context-repair, context-adjustment or free-enrichment (question Q_3). Accommodation failures can result in various forms of inappropriateness that should be examined by focussing on what Forguson calls breakdown cases (question Q_4). Finally, contextual requirements whose recognition motivates the accommodating context-changes are determined by schemas and patterns that constitute the structure of plans and intentions, where the schemas and patterns can be described in terms of rules of appropriateness constituting Austinian procedures.

References

Abbott, Barbara (2008): "Presuppositions and Common Ground". In: *Linguistics and Philosophy* 21, pp. 523–538.

Asher, Nicholas/Lascarides, Alex (2001): "Indirect Speech Acts". In: *Synthese* 128, pp. 183–228.

Asher, Nicholas/Lascarides, Alex (2003): *Logics of Conversation*. Cambridge: Cambridge University Press.

Austin, John L. (1975): *How to Do Things with Words*. Oxford: The Clarendon Press.

Bach, Kent (1987): *Thought and Reference*. Oxford: Clarendon Press.

Bach, Kent (2001): "You Don't Say?" In: *Synthese* 128, pp. 15–44.

Domaneschi, Filippo (2016): *Presuppositions and Cognitive Processes*. Basingstoke: Palgrave Macmillan.

Domaneschi, Filippo (2017): "Assertion and Presupposition in On-line Language Processing". Paper presented at a workshop *Assertion, Norms and Effects*, Trieste, 13–14 November 2017.

Forguson, L.W. (1973): "Locutionary and Illocutionary Acts". In: Isaiah Berlin (Ed.): *Essays on J.L. Austin*. Oxford: The Clarendon Press, pp. 160–185.

Gauker, Christopher (1998): "What is a Context of Utterance?" In: *Philosophical Studies* 91, pp. 149–172.

Gazdar, Gerald (1981): "Speech Act Assignment". In: Aravind K. Joshi/Bonnie L. Webber/Ivan A. Sag (Eds.): *Elements of Discourse Understanding*. Cambridge: Cambridge University Press, pp. 64–83.

Grice, Paul H. (1989): *Studies in the Way of Words*. Cambridge, Mass./London, England: Harvard University Press.

Geurts, Bart/Beaver, David I./Maier, Emar (2016): "Discourse Representation Theory". In: Edward N. Zalta (Ed.): *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). Available at: https://plato.stanford.edu/archives/spr2016/entries/discourse-representation-theory

Heim, Irene/Kratzer, Angelika (1998): *Semantics in Generative Grammar*. Oxford: Oxford University Press.

Kölbel, Max (2011): "Conversational Score, Assertion, and Testimony". In: Jessica Brown/Herman Cappelen (Eds.): *Assertion. New Philosophical Essays*. Oxford: Oxford University Press, pp. 49–77.

Kripke, Saul (1990): "Presupposition and Anaphora. Remarks on the Formulation of the Projection Problem". In: *Linguistic Inquiry* 40(3), pp. 367–386.

Langton, Rae (2015): "How to Get a Norm from a Speech Act". In: *The Amherst Lecture in Philosophy* 10: 1–33. Available at: http://www.amherstlecture.org/langton2015/

Langton, Rae (forthcoming a): *Accommodating Injustice: The John Locke Lectures 2015*. Oxford: Oxford University Press.

Langton, Rae (forthcoming b): "Blocking as Counter-Speech". In: Daniel Fogal/Daniel W. Harris/Matt Moss (Eds.): *New Work on Speech Acts*, New York:Oxford University Press.

Lepore, Ernie/Stone, Matthew (2015): *Imagination and Convention. Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Lewis, David (1979): "Scorekeeping in a Language Game". In: *Journal of Philosophical Logic* 8, pp. 339–359

Roberts, Craige (2015): "Accommodation in a Language Game". In: Barry Loewer/Jonathan Schaffer (Eds.): *The Blackwell Companion to David Lewis*. Oxford: Blackwell, pp. 345–366.

Sbisà, Marina (2002): "Speech Acts in Context". In: *Language & Communication* 22, pp. 421–436.

Sbisà, Marina (2009): "Uptake and Conventionality in Illocution". In: *Lodz Papers in Pragmatics* 5(1), pp. 33–52.

Sbisà, Marina (forthcoming): "Varieties of Speech Act Norms". In: Maciej Witek/Iwona Witczak-Plisiecka (Eds.): *Normativity and Variety of Speech Actions*. Leiden: Brill (*Poznań Studies in the Philosophy of the Sciences and the Humanities*).

Searle, John R. (1969): *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, Mass.: Cambridge University Press.

Simons, Mandy (2001): "On the Conversational Basis of Some Presuppositions". In: Rachel Hastings/Brendan Jackson/Zsofia Zvolenszky (Eds.): *Semantics and Linguistics Theory* 11. Ithaca, NY: Cornell University, pp. 431–448.

Simons, Mandy (2003): "Presupposition and Accommodation. Understanding the Stalnakerian Picture". In: *Philosophical Studies* 112, pp. 251–278.

Simons, Mandy (2013): "On the Conversational Basis of Some Presuppositions". In: Alessandro Capone/Franco Piparo/Marco Carapezza (Eds.): *Perspectives on Linguistic Pragmatics*. Cham, Heidelberg/New York/Dordrecht/London: Springer, pp. 329–348.

Stalnaker, Robert (1998): "On the Representation of Context". In: *Journal of Logic, Language and Information* 7(1), pp. 3–19.

Stalnaker, Robert (2002): "Common Ground". In: *Linguistics and Philosophy* 25, pp. 701–721.

Stalnaker, Robert (2014): *Context*. Oxford: Oxford University Press.

Thomason, Richmond H. (1990): "Accommodation, Meaning and Implicature". In: Philip R. Cohen/Jerry Morgan /Martha E. Pollack (Eds.): *Intentions in Communication*. Cambridge, MA: MIT Press, pp. 325–363.

Thomason, Richmond. H./Stone, Matthew /DeVault, David (2006): "Enlightened Update: A Computational Architecture for Presupposition and Other Pragmatic Phenomena". Target Manuscript for Workshop on Presupposition Accommodation at the Ohio State University.

von Fintel, Kai (2008): "What is Presupposition Accommodation, Again?" In: *Philosophical Perspectives* 22, pp. 137–170.

Wilson, Deirdre/Matsui, Tomoko (2012): "Recent Approaches to Bridging: Truth, Coherence, Relevance". In: Deirdre Wilson/Dan Sperber, *Meaning and Relevance*. Cambridge: Cambridge University Press, pp. 187–209.

Witek, Maciej (2013): "How to Establish Authority with Words: Imperative Utterances and Presupposition Accommodation". In: Anna Brożek/Jacek Jadacki/Berislav Žarnić (Eds.): *Theory of Imperatives from Different Points of View* 2. Warszawa: Semper (*Logic, Methodology and Philosophy of Science at Warsaw University*, Vol. 7), pp. 145–157.

Witek, Maciej (2015a): "Linguistic Underdeterminacy: A View From Speech Act Theory". In: *Journal of Pragmatics* 76, pp. 15–29.

Witek, Maciej (2015b): "Mechanisms of Illocutionary Games". In: *Language & Communication* 42, pp. 11–22.

Witek, Maciej (2016): "Convention and Accommodation". In: *Polish Journal of Philosophy* 10(1), pp. 99–115.

Leopold Hess

Expressive Meanings and Expressive Commitments. A Case of Meaning as Use

Abstract: Expressives, i.e. words such as 'damn' or 'bastard', seem to convey a specific kind of content, different from, or on top of, "regular" descriptive meaning. Following the seminal work of Chris Potts (2005), the meaning of expressives is often conceptualized within a two-dimensional semantic framework, in which descriptive and expressive contents are separated as a result of special rules of semantic composition (cf. Gutzmann 2015). This approach is successful in accounting for some interesting semantic properties of expressives, e.g. their projective behavior, and has also been extended to other classes of expressions, such as racial slurs or honorifics. However, it does not offer any actual insight into the nature of expressive meaning (the two-dimensional formalism operates on dummy values, independently of what they may stand in for).

The present paper offers an alternative, pragmatic account of expressives based on the observations that expressive meanings seem to involve directly the speaker (her states, emotions or attitudes) rather than just abstract (e.g. truth-conditional) contents, and that the utterer of an expressive is responsible for the choice of loaded, often taboo vocabulary. The account is developed in a commitment-based scorekeeping model of discourse (inspired by Lewis 1979 and Brandom 1994), in which hearers interpret speakers' utterances by attributing commitments to them. Besides assertoric commitments (and potentially other kinds), expressive commitments can be distinguished. These are commitments to the appropriateness or applicability of a given expression, which also can be attributed to speakers based on their utterances (separately from assertoric commitments).

What characterizes expressives as a lexical class is that they always raise the issue of speaker's expressive commitment. In short, expressive meanings are commitments to the appropriateness of strongly charged (often vulgar or taboo) vocabulary – which, in turn, can signal a speaker's heightened emotional state, negative attitude etc.

Keywords: expressives, meaning as use, multi-dimensional semantics, commitment, expressive commitments, scorekeeping

1 Introduction

The topic of my paper is the nature of expressive meaning, which I will construe narrowly as the meaning of expressives (although, as will be suggested later, the category of expressive meanings may be fruitfully extended to other linguistic items).[1] Expressives are strongly charged, emotional, often taboo (or highly colloquial) words such as fuck, shit, damn, bastard. As these few examples already show, expressives can belong to any grammatical category – they can serve as nouns, adjectives, verbs, interjections, etc. Their main function seems to be to manifest a negative (less often positive) attitude towards a situation or a heightened emotional state. They typically do not have any descriptive content, or if they do (e.g. 'fuck' used as a verb to refer to sex), it is easily detachable from an expressive component ('fuck', when used to refer to sexual activity, can be replaced with 'have sex' with arguably no change to truth conditions).[2]

I will begin with a brief discussion of the main properties of expressives in section 2. In section 3, I will discuss an influential account of expressive meanings within the framework of multi-dimensional semantics as developed by Chris Potts. While formally cogent and offering interesting extensions to other classes of expressions, the semantic approach is severely limited – mostly importantly, it offers no insight into what expressive meanings as such are, representing them only with dummy values. In the brief section 4, I will present some intuitive support for a different, more pragmatically oriented approach. It focuses on the speaker's choice to use a marked, "risky" expression and the resulting responsibility. In section 5, I will show how this notion can be modelled within a commitment attribution framework based on Lewisian conversational scorekeeping. In section 6, I briefly conclude.

The semantic approach to expressive meanings is inspired partly by an idea of David Kaplan, who envisaged extending formal semantic methods to capture a notion of *meaning as use* (which, in more or less vague forms, has been a fixture of 20th century philosophy of language). In proposing an alternative account of expressive meanings, I will also suggest another way of thinking about meaning as use – one that is in a sense a more direct and stricter application of this idea.

---

2 The boundaries of the category of expressives are fuzzy. The characterization presented so far does not clearly determine whether *idiotic, awesome,* or *totally* are expressives. My strategy here is not to engage such problems, but to focus on the unambiguous core examples. Non-prototypical instances may share relevant properties to some extent, and my arguments may apply to them partially.

2 Properties of expressive meanings

Christopher Potts (2007) names the following properties of expressives:[3]
– Independence
– Nondisplaceability
– Perspective dependence
– Immediacy
– Descriptive ineffability
– Repeatability

Of these, the first three are most important for my present purposes. "Independence" describes the fact that expressives contribute a dimension of meaning which is independent of the main content of the utterances in which they appear. They serve only as a kind of comment on or supplement to the main – at-issue, in semanticist jargon – content and can be removed or replaced with neutral, unmarked counterparts without any change to truth conditions.[4]

"Nondisplaceability" means that expressives always predicate something of the context of the actual utterance – therefore, their content cannot be displaced even in contexts such as speech and attitude reports, modal or conditionalized statements, reports of past events, etc.

"Perspective dependence" describes the fact that expressive content is always evaluated from a concrete perspective. In general, the perspective is the speaker's, but under some conditions it can be someone else's.

The other properties listed by Potts are less important, and in fact less tangible. The precise content of expressives is difficult to articulate ("descriptive ineffability"); it is non-negotiable and, in a way, directly imposed on the speech situation ("immediacy"); expressives can be repeated to strengthen their effect ("repeatability"). I will not discuss these features at present.

The following two examples illustrate Independence and Nondisplaceability.

---

3 My focus is on expressive meanings, which include the meanings of expressives, but may be a broader category (honorifics, slurs, formal pronouns etc. may convey expressive meanings, but they are not expressives).

4 It is worth noting that being not-at-issue and having non-truth-conditional content are two distinct properties, even if this distinction is obscured in Potts' original treatment of expressives. Gutzmann (2015) offers a refinement of Potts' theory in which the two are disentangled. However, the distinction will not be relevant to any aspect of my discussion here, and I will treat the two notions as equivalent.

(1) We bought a new electric clothes dryer [...] Nowhere did it say that the **damn** thing didn't come with an electric plug!

(2) If that **bastard** Kaplan got promoted, then the Regents acted foolishly (Potts 2005).

In both cases the content of the expressive makes no contribution to the at-issue content of the sentence. What is conveyed is clearly a negative attitude or emotional state of the speaker, but this is only a comment, in a way, on what the main content of the utterance is. On the other hand, the embedding of the expressive in a report and under negation in (1) or in an antecedent of a conditional in (2) makes no difference to what it communicates. Expressive content is not conditionalized, negated, or evaluated with respect to the context of an original speech act, but should be interpreted with respect to the actual speech situation.

In (1) and (2) the negative attitudes expressed by "damn" and "bastard" are clearly those of the speaker her- or himself. This is a typical situation – there is a (very strong) default interpretation of expressives as conveying the speaker's perspective. However, as mentioned before, occasionally the perspective may be someone else's[5]. Consider (3):

(3) My father screamed that he would never allow me to marry that **bastard** Webster (Amaral *et al.*, 2007).

The most plausible interpretation of (3) is clearly that it is not the speaker's negative attitude toward Webster that is expressed (as he or she apparently intends to marry Webster), but that of the speaker's father. The expressive is embedded under a speech report, but we already saw in (1) that this is not enough by itself to evoke a non-speaker-oriented reading of an expressive. Higher-level pragmatic factors are clearly at play here.

I assume that the three basic properties of Independence, Nondisplaceability, and Perspective dependence are the minimal explananda for a successful account of expressive meanings.

---

5 It is very difficult to define precisely what the conditions are under which non-speaker-oriented readings of expressives are possible. They do not seem to have anything to do with syntactic or semantic embedding (cf. "nondisplaceability" above). See Potts (2007), Harris and Potts (2009) and Hess (2018) for a discussion. I will briefly return to this point in section 3.2.

3 Two-dimensional semantics for expressives

3.1 Potts' Logic of Conventional Implicature

In an unpublished but highly influential lecture, David Kaplan[6] proposed to extend the methods of formal semantics to linguistic categories that had so far been largely ignored by semanticists – including expressives.[7] The way to do that was by attending to the rules of use of relevant expressions ("the right *sort* of rules of use", Kaplan insisted). Stating the rules of use in a formal setting, rather than providing a translation or a denotation, could constitute an account of their meaning, because for the relevant kinds of expression "meaning *is* use". The inspiration for that project was, of course, Kaplan's own theory of indexicals:

> Take for example the first person pronoun *I*: It seems fruitless to ask what the first person pronoun means; as Frege said, it seems to mean different things on different occasions of use. But the question, "What are the conditions under which the first person pronoun would be correctly used?" quickly yields a good answer, namely: to refer to the person who uses it (Kaplan 1999, p. 3).

For words such as *damn* or *bastard* it could be said, analogously, that they are correctly used to express the speaker's negative attitude. And just as with indexicals, Kaplan argued, such rules could be integrated into an account of truth and validity, so that the two kinds of meaning present in a sentence like *That bastard Kaplan got promoted* could be unified in a single semantic theory.

Kaplan himself stopped short of developing these suggestions into an actual formal theory, but this was done by Christopher Potts. Potts (2005) elaborates on Kaplan's core idea of meaning-as-use in a formal semantic setting and connects it with the concept of *conventional implicature*. The latter comes from Grice (1975) – a conventional implicature (CI) is a part of the meaning of a linguistic expression that does not contribute to the truth conditions of a sentence (does not belong to *what is said*), but it is conveyed as a matter of convention (and is not calculated in context

---

6 Kaplan 1999; the lecture is known under the title *The Meaning of Ouch and Oops: Explorations in the Theory of Meaning as Use*. I quote from a transcript prepared by Elizabeth Coppock and available online: http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf (accessed on October 10th, 2017).

7 Kaplan also mentioned interjections such as *ouch* or *oops*, diminutives, nicknames, greetings, ethnic slurs, etc. The common denominator is that the meanings of these words does not seem to be of the same descriptive type as *fortnight* or *feral*. Whether or not they can all be identified with expressive meanings in the sense I am discussing in this paper is an open question.

like a conversational implicature). Grice's original examples of CI-triggers were words like *but* or *therefore*. For instance, in (4), what is said (at-issue content) is that Donovan is poor and that he is a happy, while the conventional implicature is something like "there is a contrast between being poor and being happy".

(4) Donovan is poor but happy.

According to Grice, *but* has the same truth-conditional import as *and*, but in addition triggers the implicature of contrast. Bach (1999) criticized this approach, arguing that the supposed implicatures of words like *but* or *therefore* can be displaced, e.g., in speech reports. Potts in turn does not defend the Gricean candidates for CI-triggers, but offers new ones: appositives and expressives. Consider example (5).

(5) Lance Armstrong, **a cyclist**, survived cancer.

Here, the at-issue/truth-conditional content is that Lance Armstrong survived cancer. The appositive (in boldface) conveys a CI that Lance Armstrong is a cyclist. The CI is by default speaker-oriented (it is not displaced, e.g., in speech or attitude reports), which makes it possible for sentences such as (6) to be felicitous.

(6) Jack thinks that Lance Armstrong, a cyclist, is a Formula 1 driver.

The stipulation that appositives trigger conventional implicatures is therefore immune to Bach's criticism. The same goes for expressives, the non-displaceability of which was already mentioned. (I will not discuss appositives further.)

Example (7) presents the basic schema of Potts' CI theory of expressives. BAD is a dummy predicate that stands in for a representation of the speaker's negative attitude which is conveyed – through being conventionally implicated – by the expressive "bastard". I will come back to this point shortly.

(7) That bastard Kaplan got promoted.
     At-issue: Kaplan got promoted.
     CI: BAD(Kaplan)

According to Potts, the two levels of content – at issue and CI – are independent and undergo transformations separately. Accordingly, any displacement or embedding can only target the at-issue part of the content. Thus in (8) the conventional implicature remains unaffected (BAD still represents the speaker's attitude, not Jack's):

(8) Jack said that that bastard Kaplan got promoted.
   At-issue: Jack said that Kaplan got promoted.
   CI: BAD(Kaplan)
   (Potts initially ignored examples of perspective shifts such as (3).)

This analysis is developed formally in a two-dimensional semantics in which at-issue/descriptive and CI/expressive contents[8] are processed separately at every level of the theory: there are separate at-issue and expressive types, CI-functional application rules and special interpretation rules for objects that are combinations of descriptive and expressive contents. I will not describe the details of Pott's formalism, as nothing in my discussion hinges on them. Just as an illustration, to give the reader some sense of what it looks like, here is a toy example. It shows a fragment of a *semantic parsetree*, which is a structure resulting from type-driven translation of a natural language sentence (where combinatoric semantic rules can be treated as tree-admissibility conditions) and which can be model-theoretically interpreted.[9]

(9) The damn Republicans are aggressively cutting taxes (Potts 2005).

$$\textbf{republican} : \langle e^a, t^a \rangle$$

$$\bullet$$

$$\textbf{damn(republican)} : t^c$$

$$\textbf{republican} : \langle e^a, t^a \rangle \qquad \textbf{damn} : \langle \langle e^a, t^a \rangle, t^c \rangle$$
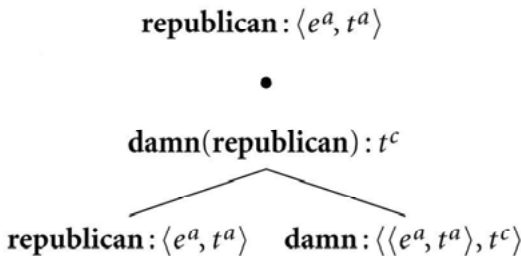
Fig. 1: Partial semantic parsetree for damn Republicans (Potts 2005, p. 166).

---

8 As I will not be interested in appositives or any other potential examples of conventional implicature, I will identify CI contents with expressive contents for now.

9 Semantic parsetrees, therefore, belong to a semantic translation language, which turns out to be indispensable in Potts' theory, contrary to the belief of many semanticists that direct interpretation is to be preferred.

The diagram presents a part of the semantic parsetree for (9) (a local tree) corresponding to the composition and translation of "damn Republicans". The a-superscripted and c-superscripted types are separate types (at-issue and CI/expressive, respectively). The expressive "damn" here has a type which takes as input a (descriptive) predicate and returns an expressive t-type object. However, the special rule for functional application where CIs are involved is, importantly, not resource-sensitive, so "damn" returns its argument as well. Thus, the meaning of "damn" passes both its argument and a resulting CI proposition up to its mother node. The bullet (•) represents the separation of the at-issue and CI contents, which from then on are processed independently (the CI proposition is in fact passed unchanged to the top of the parsetree and interpreted together with the at-issue proposition).

How is the part of the tree below the bullet (•) interpreted, or in other words, what is the CI proposition? As suggested already by Kaplan the semantics for *damn* simply states that it expresses a speaker's negative attitude. The meaning of the CI proposition in (9) is therefore that the speaker holds a negative attitude towards Republicans. Such meanings can be incorporated in the formal theory by providing lexical entries for expressives like the following:

$$\left\{\begin{array}{c} damn \\ bloody \\ \vdots \\ fucking \end{array}\right\} \rightsquigarrow \lambda X.\,\mathbf{bad}(^{\cap}X):\langle\langle\tau^a, t^a\rangle, t^c\rangle$$

Fig. 2: Lexical entries for expressives (Potts 2005, p. 167).

The function **bad** is the function that says that the speaker has a negative attitude towards X.[10] (The $^{\cap}$ symbol is a nominalizing type shifter.) When this meaning for "damn" is applied to the meaning of "Republicans" we get the desired proposition, one that holds (is expressively correct) if and only if the speaker has a negative attitude towards Republicans.

---

10 Considering that expressives can be used not only to convey negative attitudes and emotions, but also positive ones, or simply states of agitation or excitement, Potts suggests an even more general translation: "the speaker is in a heightened emotional state about X".

Potts' Logic of Conventional Implicature, as he called the system, is a powerful formal tool which makes it possible to carry out Kaplan's postulate of extending semantic methods to new kinds of content. The treatment of expressives within LCI straightforwardly accounts for two of their most salient properties: Independence and Nondisplaceability (I will discuss Perspective dependence later), because the CI propositions resulting from the admission of expressive meanings in a semantic parsetree are always interpreted at the top node (thus projecting from all embeddings) and separately from the at-issue content.

Slightly modified versions of Potts' LCI have been applied to other linguistic categories that share some properties with expressives of the *damn* and *bastard* sort. For example, one plausible theory of *slurs* (i.e. racial, ethnic or sexual pejoratives) treats them as *mixed expressives* – words that carry both a descriptive and an expressive meaning. They would thus contribute to both dimensions of meaning in the Pottsian system. The lexical entry in (10) is a toy analysis of a (somewhat archaic) slur for Germans, as suggested by Eric McCready (2010). The bullet (•) separates two kinds of content, as usual. The × symbol indicates that the type of the expression is compound: it has the (at-issue) predicate type and also the expressive proposition type (here indicated by s-superscript).

(10) $[[\text{Kraut}]] = \lambda x.\text{German}(x) \bullet \text{bad}(\cap\text{German}) : <e, t>^a \times t^s$ (McCready 2010).

This meaning for *Kraut* indicates that when a speaker uses this slur, she is referring to a German individual (as if she just said "German") and at the same time expressing a negative attitude towards all Germans as such.

McCready offers a similar treatment for Japanese *honorifics*: expressions (especially verbs) that combine descriptive content with a meaning that conveys the speaker's reverent attitude towards the verb's subject or another agent. An example of his analysis can be seen in (11) and (12).

(11) $[[\text{irassharu}]] = \lambda x.\text{come}(x) \bullet \text{honor}(x) : <e, t>^a \times <e, t>^s$
(12) Yamada-sensei-ga irasshaimasi-ta
      Y-teacher.NOM came.HON.PST
      'Teacher Yamada came. (And I honor him.)' (McCready 2010)

Other linguistic categories that have some aspect of expressive meaning could be treated similarly. Potts (2007) proposes an expressive-meaning account of formal and familiar pronouns (e.g. the French *vous* and *tu*). Daniel Gutzmann (2015) develops an extended version of the LCI system that covers categories that are not

superficially similar to expressives, like sentence mood. He also refines the treatment of expressives proper by distinguishing several types based on the relation between the descriptive and the expressive meaning dimension.[11]

The Kaplan-Potts line of thinking about expressives is clearly quite successful in answering some important questions and can be productively applied to other areas. However, it also has many limitations, which I discuss in the next section. In the end, I argue, even if formally sound, it is not philosophically satisfactory as a theory of expressive meaning.

3.2 Limits of the semantic theory

One problem with Potts' two-dimensional semantic theory of expressive meanings – which I do not consider to be very serious, but it is worth bringing up – is that it is not, strictly speaking, compositional. For one, the CI functional application rule is not resource sensitive (the meaning of *damn* passes up its argument unchanged, along with the expressive proposition) – but Potts argues quite convincingly that resource-sensitivity is not essential. The flipside of this is, however, that a plausible way of defining compositionality, as a requirement that the meaning of an expression be dependent only on the meanings of its direct components, is not satisfied here.[12] For the CI proposition needs to be passed to the top node of the parsetree to be interpreted. That is, the meaning of the top node expression, e.g. the meaning of the whole sentence *The damn Republicans are aggressively cutting taxes*, is a compound of the meaning of its direct components and the meaning of *damn*'s expressive proposition. The (at-issue) sentence meaning, after all, is composed using the same rules as it would if it did not contain any expressives (basically applying the meaning of the verb phrase to the meaning of the noun phrase), as the latter's meaning is independent and not at-issue. And nothing in those rules deals with the content that comes after the bullet.

This way of framing the problem already suggests a solution – amend all rules of composition so that they take into account both at-issue and expressive meanings of the expressions to be composed. This is, in fact, the way taken by Gutzmann (2015) in his extension of LCI. The extended system is formally precise and compositional – however, it requires that all expressions of the language

---

11 Mixed expressives such as slurs and honorifics are of a different kind than ones that only contribute to one dimension while leaving the descriptive content in place; it is also possible for an expressive meaning to "shift" a whole descriptive proposition to the expressive dimension (like an exclamative).

12 This point was first raised by Barker et al. (2010)

have two-dimensional meanings. This is easily handled technically by adding a lexicon-semantics interface rule which adds dimensions with trivial meanings to all expressions (e.g. a descriptive word such as *dog* would get an identity function as its expressive meaning, so that it contributes nothing to the expressive dimension, but can be composed multi-dimensionally).[13]

Because of the availability of solutions such as Gutzmann's, I do not take the non-compositionality of Potts' original two-dimensional semantics to be a very serious issue, but one has to notice that the cost of the solution is the loss of the intuitive simplicity of Pott's basic idea of the logic of conventional implicature. It is easy enough to see why the meanings of words such as *damn* or *bastard* should be computed as separate but integrated parts of the meaning of the sentences in which they appear. It is much more difficult to see that all sentences should have multi-dimensional meanings, and that all expressions should contribute to all the dimensions (even if, or maybe especially if, most of these contributions and the resulting contents are trivial). As an account of expressive meanings, this would require much more independent justification to be plausible.[14]

Another, slightly more serious problem with the Pottsian approach to expressive meaning concerns the relation between the argument of the **bad** function and the object of the speaker's attitude. Consider the following two examples:

(13) **My damn car** won't start again!
(14) I can't remember where I parked **my damn car**!

In the semantic parsetrees for both (13) and (14) the meaning of "damn" would be applied to the meaning of "car" (I ignore any interaction with the possessive pronoun) producing the CI proposition that the speaker holds a negative attitude towards the car. This proposition undergoes no further transformation and it is interpreted alongside the at-issue proposition once the whole tree is composed. But the result is counterintuitive – the most natural way to understand (14) is that the speaker is not actually upset with the car (as they clearly are in (13)), but rather with the whole situation as such.

---

13 In fact, in Gutzmann's extended system, all expressions have three dimensions of meaning to accommodate different relations between different types of content, but the general idea can be illustrated with just two dimensions, as I do here.

14 Note that Gutzmann does provide such justification, as his is a theory of "use-conditional" meanings construed much more broadly – expressives are just one among many disparate applications. However, this does not in any way help with what I consider to be more serious deficiencies of the semantic approach to expressives, which I discuss below.

There is no principled way of resolving this issue within Potts' theory or its extensions. Formal solutions are of course easily available – one way to get the intended reading of (14) would be to posit a type shift, so that the meaning of "damn" would be applied elsewhere in the tree and take the whole at-issue proposition as its argument. But (and this would probably go for any other solution, whatever the technical details) this is entirely *ad hoc*. There is nothing in the syntax of this sentence to indicate a type shift; in fact, there seems to be nothing to suggest that the syntax or semantics of "damn" in (14) are any different than in (13).

This is hardly a fatal flaw in the multi-dimensional semantics of expressives, but it does reveal the limits of this approach. The theory succeeds at integrating expressive meanings into a compositional (with caveats) formal framework, but it does very little to illuminate the ways in which they contribute to expressing what the speakers have to say.

A more serious objection against the semantic theory of expressive meaning is that it only successfully accounts for two of the three explananda that I posited as minimal requirements in section 2. Independence and Nondisplaceability are, manifestly, core properties of Conventional Implicatures in Potts' account. What about Perspective dependence? In his original theory, Potts assumed that CIs, including expressive meanings, are always speaker-oriented. Counterexamples were soon found, and many authors, including Potts himself (see Potts 2007 and Harris and Potts 2009) later argued that non-speaker-oriented readings, in which an expressive appears to convey an attitude of someone other than the actual speaker, are possible. Let me first present a few examples before I discuss the theoretical ramifications that follow from this.[15]

(15) My father screamed that he would never allow me to marry that **bastard** Webster.
(16) [Context: We know that Bob loves to do yard work and is very proud of his lawn, but also that he has a son Monty who hates to do yard chores. So Bob could say:]
(17) Well, Monty said to me this very morning that he hates to mow the **friggin** lawn. (Amaral *et al.* 2007)

---

15 See Hess (2018) for a detailed discussion of the perspectival dependence of expressives and the relevant literature, as well as an extended criticism of semantic solutions. The suggestion that a *pragmatic* mechanism of perspective shift is involved was first made by Harris and Potts (2009).

(18)   My classmate Sheila said that her history professor gave her a low grade. The **jerk** always favors long papers (Harris and Potts 2009).

Examples (15) (repeating (3) from section 1) and (16) come from a critical review of Potts' *Logic of Conventional Implicature* (Amaral et al. 2007). The most natural reading in both cases attributes the negative attitude conveyed by the expressive to a non-speaker agent:[16] the father in (15) and Monty in (16). This is suggested by the context (explicitly provided in (16) and easily inferred in (15)) which contains the information that the actual speaker's attitude towards Webster and yardwork, respectively, is very positive and therefore incongruous with the use of a negative expressive. This incongruence is easily resolved by taking the speaker to be some-how mimicking (whether by directly reporting or indirectly representing) the ac-tual words of the subject.

One could notice that in both (15) and (16) the expressive is embedded in a speech report, so the apparent shift in perspective can be a function of that. An analogy would be the way *tasty* shifts to a non-speaker-oriented reading in (18).

(19)   John said his **tasty** lunch was very expensive.

On a strongly preferred reading of this sentence, it is John's, not the speaker's taste that is relevant to the evaluation of the food. However, as we've already seen in (1), embedding in a speech report does not always shift the interpretation of an expressive – and the property of Nondisplaceability implies there should be no connection between overt embedding and perspective shifts.

Example (17) demonstrates the converse point. It comes from an experi-mental study conducted by Harris and Potts (2009) in which respondents were confronted with items like this one and were asked whose opinion it is that the professor is a jerk, the speaker's or Sheila's. In this and many other variants of the example (with different situations and different expressives), as many as 45% respondents answered that it is the subject's (i.e. Sheila's in this instance) atti-tude that is conveyed by the expressive. Many respondents were also undecided, which may indicate that both answers seemed likely to them. Thus, while a speaker-oriented reading of "jerk" in (17) is possible, the non-speaker-oriented reading cannot be discounted. And the most important point is that the expres-sive here is not overtly embedded in a speech report.

---

16   That is, an agent other than the *actual* speaker (which may be, and in the relevant cases very often is, a reported speaker).

It appears from these and many other examples that can easily be found or constructed that, while there is a strong bias towards a default speaker-oriented interpretation of expressives, perspective shifts to other individuals are possible in suitable contexts.[17] How can the semantic theory of expressive content account for this?

There are several variants of the answer,[18] but they all come down to positing an indexical analysis – which *prima facie* seems like the natural choice. The central idea is simple: there is a special parameter (like a judge or evaluator) in the context which accounts for the orientation of expressive content. By default, this parameter is set to the actual speaker, but in some circumstances it can be set to another salient agent, giving providing a non-speaker-oriented interpretation.

Without going into the details of possible formal implementations, let me quickly point out the problems with this solution. Firstly, and most importantly, it is explanatorily idle. There are no syntactic or semantic conditions definable by the theory that would engender a shift of the relevant contextual parameter – speech reports sometimes seem to facilitate a perspective shift of the expressive meaning, but not always; and sometimes they are not needed. The indexical analysis cannot predict the shifts in either case. And it is easy to see that no other potential predictor works better (again, this is a corollary of Nondisplaceability). This problem is similar to the one we have noticed already regarding the object of the attitude manifested by an expressive (see examples (13) and (14)) – technical devices are easily available that make it possible to incorporate any interpretation of expressives into the formal semantic apparatus, but they are always entirely *ad hoc*. This strongly suggests that more interesting answers lie elsewhere, beyond the purview of a formal semantic theory.

Furthermore, besides theoretical futility, the empirical adequacy of the indexical analysis with respect to perspective shifts of expressive meaning can be called into question. For the distribution of shifted expressives shows no apparent regularity and does not seem to resemble the distribution of other shiftable categories to which similar formal approaches have been plausibly applied. Take for instance predicates of taste, such as *tasty* – it is on Lasersohn's (2005) account of these, which involved a shifting judge-parameter, that Potts (2007) modeled his indexical treatment of expressives. We have already seen in (18) that the meaning of predicates of taste normally shifts to a subject-orientation in speech reports (and to a hearer-orientation in questions, etc.), while this is not in general

---

17 What those contexts are exactly is not easy to define, but they most often seem to be narratives.

18 See Potts (2007) and Schlenker (2007) for two main options.

true of expressives. To underscore the contrast, let us consider the following minimal pair.

(20) Jim said that he ate some **tasty** Brussels sprouts again.
(21) Jim said that he ate some **damn** Brussels sprouts again (Hess 2018).

The predictable shift of perspective occurs in (19) – the most salient reading is clearly that it is Jim who finds brussel sprouts tasty. Without some additional cues, however, no such shift seems to occur in (20). It is the speaker, not Jim, who appears to hold a negative attitude (towards the food item, or maybe Jim's eating habits, etc.). The same will hold for any other kinds of context where the meanings of predicates of taste easily or obligatorily shift – expressive meanings will by default remain speaker-oriented.[19]

    I have mentioned three properties which I assume to be minimal explananda for a theory of expressive meaning: Independence, Nondisplaceability, and Perspective dependence. A semantic theory of the kind first suggested by Kaplan and then developed by Potts and others provides a plausible and quite elegant account of the first two properties. It fares much worse when it comes to the third one – even if it can offer a formally cogent account of perspective shifts with expressives, it does not in fact explain anything.

    Moreover, I have repeatedly stated that these are *minimal* explananda – and I want to suggest that even if they were satisfactorily dealt with, it would not be sufficient. For expressive meanings seem to be special, and very unlike the meanings of such words as *dog* or *every*, or even *tasty* – they *express* emotions and attitudes rather than describe objects and situations; they have a certain intensity or immediacy and reveal the speaker's thoughts in a way that is much more direct than in the case of simple assertions. The semantic theory is – by design – not concerned with those aspects, and in effect does not illuminate the nature of expressive meanings in any philosophically interesting way. We do not learn much, after all, about the meaning of *damn* by looking at the dummy function **bad**.

    In light of this discussion, I wish to offer an alternative approach to the semantic theory of expressive meanings. Because I do not claim that the semantic

---

19 Schlenker's (2007) account of expressive perspective shifts is modelled on his theory of "shifting indexicals". (In some languages personal pronouns may shift, e.g. *I* in speech reports may refer to the subject, not the speaker.) Shifting indexicals, however, have very regular distributions; it is relatively easy to define in what kind of conditions their shifts are possible or obligatory – and these conditions can be formalized. Again, there is no similarity here with expressives.

theory is, strictly speaking, incorrect (nor do I offer a formal treatment of expressives), my pragmatic account may be treated as complimentary to it – but I do offer alternative answers to the question why expressives are characterized by the three main properties, and argue that their meaning is different *in kind* (not only in the way it is composed) from descriptive meanings.

In section 4 I briefly present some examples and reflections to motivate a certain way of thinking about expressives, which I will then develop into a more explicit account in section 5.

4 Speaker's choice

The first point that I want to bring up to suggest a different approach to expressive meanings is the way in which one can respond to an expressive-containing utterance, if one does not agree with or share the speaker's attitude. Expressives cannot be challenged or negated directly, but only through what may be called a *meta-linguistic challenge*. Consider the variants in (21):

(22) A: I'm not going to mow the **friggin'** lawn.
     a. B: **#** The lawn is not friggin'. / **#** You don't hate mowing the lawn.
     b. B: Watch your language, son! / That's not a word we use in this house.

The answers in (21a.) – either targeting the expressive directly or challenging the attitude it supposedly reflects – are not felicitous. They seem to miss the point somehow. The answers in (21b.), however, are perfectly natural and exactly what one would expect from, say, a parent confronted with their teenage son's outburst. Those answers do not target the content of "friggin'" directly, but rather challenge the way in which the speaker A chose to express himself.[20]

Note that this is different than meta-linguistic negation in a traditional sense,[21] as in *The king of France isn't bald – there is no king of France*. Existential presuppositions, however, are not the only possible targets of meta-linguistic ne-

---

20 An anonymous reviewer suggested that an answer like "Come on! I know you don't hate mowing the lawn." would also work. I am not sure if it is the expressive that is challenged here. But in any case this is also a meta-linguistic challenge in an important way: it targets the sincerity conditions of the utterance (Monty's attitude).
21 Cf. Horn (1985).

gation. Sometimes it can have more to do with differing perspectives on a situation (*The glass isn't half full – it's half empty*) or the choice of near-synonyms (*I'm not happy – I'm ecstatic!*). And something like this is possible with expressives, too:

(23)   That BASTARD Kaplan didn't get promoted – your GOOD FRIEND Kaplan did![22]

What I mean by *meta-linguistic challenge* is a broader category comprising both cases of meta-linguistic negation in a standard sense like (22) and responses such as (21b), and anything in between. Two important points here are that, one, these are ways of challenging the way in which certain content is expressed (refusal to mow the lawn, Kaplan's promotion) rather than the content of some additional items (like the existence of the king of France); and two, that it is a characteristic feature of expressives that they can only be challenged felicitously like this.

   The importance of noting the ways in which expressives can be negated or challenged lies precisely in the *meta-linguistic* aspect. It is this aspect that is challenged in a felicitous response, because the speaker's choice of expression is intrinsically related to the very meaning of expressives. Consider the two ways I could express my dissatisfaction with the semantic account of expressives (note that I'm not actually claiming either of these things):

(24)   This theory is completely wrong.
(25)   This theory is **bullshit**.

I assume that (23) is more or less the strongest way of saying that the theory in question is bad without actually swearing (if needed you can substitute "utterly misguided", or "absolutely erroneous", etc.). It does not seem like (24) adds anything to that content-wise. And yet if I wrote that, or better yet, said it during a conference presentation, it would have a stronger effect – the audience would surely take note that I used language that is not normally considered acceptable in such a setting, and therefore could be entitled to infer that I consider the theory so bad that it warrants this kind of strong language.[23] (Both aspects, the breach of decorum and the warranting of strong language, will be central to my positive account in the next section.)

---

22 *That bastard Kaplan didn't get promoted – he's not a bastard* is bad, indicating that expressive meanings are not presuppositions, and that there are important differences between types of meta-linguistic negation.

23 This is not to say that the expressive meaning is an intensifier – "completely" plays this role in (23) to a very different effect.

We see in this example that the speaker's choice itself is in some way a vehicle of meaning. This is why there is such a strong bias for speaker-oriented interpretations of expressives – the speaker is the one responsible for the choice of words, even when those words appear, for instance, in speech reports or deeply embedded contexts. Let me illustrate this point with a lovely quote from *The Wire*:

(26) MCNULTY: You know something? My ex-wife, the way she acts sometimes, the way she deals with shit... You would think a less enlightened man than myself, a cruder man than myself, a man less sensitized to the qualities and charms and value of women – a man like that, not me, but a man like that, he just might call her a **cunt**.
GREGGS: You just called the mother of your children a cunt ("The Pager", *The Wire*, HBO).

In his utterance, McNulty goes to great lengths to distance himself from the expressive he will cite as something "a cruder man" could call his ex-wife, Elena. This is of course entirely transparent to Greggs, who calls him out on it, and rightly so. Three things should be stressed here. First is that despite all the hedging and embedding, the expressive "cunt" is correctly interpreted as conveying the attitude McNulty holds towards its referent, i.e. his ex-wife. Second, he does not simply reveal his attitude, he can be held responsible and blamed for the use of the expressive. It's not just that McNulty expresses his feelings towards the mother of his children, he actually *calls* her a "cunt". Third, Greggs does neither of these things. She reports on McNulty's use of the expressive, and despite the projective tendency of expressive content, she is not responsible for the vulgar language and, in effect, she is not interpreted as expressing her own negative attitude towards Elena.

The intuitive idea which I want to introduce through the examples in this section is as follows. Expressives are strong, marked expressions; a speaker using them may be held responsible for the choice of words, and directly challenging this choice (rather than any content conveyed) is the most important way of challenging the expressive. Moreover, it is the speaker's choice that is the vehicle of expressive meaning – roughly, by choosing an expressive the speaker signals that it adequately reflects his emotional state, even if, or rather precisely because it constitutes a breach of decorum. In the next section I will develop this idea in a theoretical framework, employing commitment attribution and Lewisian "scorekeeping in a language game".

5 Expressives on a conversational scoreboard

To preview quickly the account I offer here, I will posit that expressives are *pragmatically opaque*, in that rather than simply (transparently) conveying meaning (as descriptive items do) they focus the audience's attention on the speaker's choice of words.[24] That way they make the speaker's *expressive commitments* salient. These are to be distinguished from assertoric commitments. Attributing commitments to speakers and other agents is the core of discourse interpretation, which I will frame in Lewis's "scorekeeping" model. Attributing an expressive commitment to a speaker is how one interprets an expressive.

I will begin with a presentation of the scorekeeping framework based on Lewis' (1979) classic account of "conversational kinematics". A conversational scoreboard is an abstract object which keeps track of certain parameters that are important for the flow of conversation, such as, e.g., the standards of precision for vague terms, salience of potential referents, shared presuppositions, etc.[25] A good example is the first category mentioned. Imagine that Tom is taller than Jerry, and that one of the participants says that Jerry is tall. If no one objects, the scoreboard records that standard for the vague predicate *tall* has been set at such a level that Jerry qualifies as tall. Therefore, in the following conversation, it cannot be felicitously denied that Tom is tall as well, even though it is perfectly possible, on another occasion, to set a standard according to which he would not be.

The only items tracked on the scoreboard which I will be interested in are commitments. I will introduce expressive commitments shortly, but one of the most important parts of the scoreboard are assertoric commitments – commitments of participants of the discourse to the contents of the assertions they have made (this can be further extended to commitments relating to other speech acts). The basic principle of scorekeeping is that participants of a discourse should keep track of anything that makes a difference to what subsequent "moves" in the conversation are possible. And assertoric commitments, of course, make immediate

---

24 A helpful way of articulating this sense of opacity, suggested to me by an anonymous reviewer, is by saying that expressives give rise to a *meta-linguistic implicature* to the effect the speaker is aware of breaking a standard of linguistic decorum, and takes responsibility for it because he or she believes that it is warranted.

25 The metaphor comes from games such as baseball, where many different values are tracked throughout the game (strikes, balls, outs etc.) and the score at any given time (e.g. which team is on offense or the number of strikes on the batter) influences the eligibility and significance of further plays.

and significant difference regarding what further utterances in the conversation are allowed, felicitous, or expected.

Lewis originally posited a common scoreboard for all participants of a conversation, and there is of course a point to keeping track of those assertions to which all interlocutors have assented (to reflect the Stalnakerian Common Ground) – but there are also reasons to keep track of individual commitments of participants, and models of discourse that do this have been proposed by linguists and philosophers. For instance, Farkas and Bruce (2010) use individual discourse commitments to model the function of polar questions and reactions to assertions. Brandom (1994) develops a theory of language use in which the "kinematics" of assertoric commitments and resulting entitlements and justificatory obligations are crucial for an account of sentence meaning – this also requires keeping track of individual commitments of speakers.[26]

The basic setting I adopt here is therefore this – a scoreboard model of discourse which keeps track of participants' individual commitments (I ignore any other elements of the scoreboard). I will now introduce *expressive commitments* into this framework. While assertoric commitments are commitments to the truth of asserted contents, expressive commitments concern the way contents (of assertions or other speech acts) are expressed – they are commitments to the applicability or appropriateness of a certain term or expression. They may be salient in a given context because there are importantly different ways of referring to a given object or situation, reflecting diverse perspectives or carrying various associations and connotations. A speaker may wish to bring those perspectives and associations into focus or downplay them (or they may do that inadvertently). Consider the following examples:

(27) A: The **freedom fighters** succeeded in liberating the village.
      B: Those '**freedom fighters**' butchered its inhabitants. They're **terrorists**! (Brandom 1994)[27]
(28) John made Mary **a beergarita**.
      b. John made Mary **what he calls a beergarita**. (Harris 2014)

---

26 Unlike Farkas and Bruce, Brandom explicitly frames his commitment-tracking account in terms of conversational scorekeeping. I am following his lead here. The concept of expressive commitment also comes from Brandom, although he seems to reserve it for the use of expressions that differ in inferential significance – this seems to exclude such "trivial" or extra-linguistic connotations as are relevant in (27) or (28) below.

27 Example modified.

(29) A: Do you want a **pop**?
     B: No, but I'd love a **soda**.

In the exchange in (26) speaker A calls a group of militants "freedom fighters" –
a term with clearly positive associations – which correspond to her perspective
on the events in question (note that "liberate" is also a marked expression). The
speaker B disagrees with the assessment of what happens and calls the militants
"terrorists" – a term with a strong negative load. Both speakers may not disagree
about the facts of the situation (the militants took control over the village, they
killed many people), but they do differ in their interpretations, and because of
that they take different expressions to be appropriate. A *commitment* to the ap-
propriateness of "freedom fighters" and "terrorists", respectively, may be at-
tributed to them and recorded on the conversational scoreboard. (Quite obvi-
ously, these commitments are recorded in the individual columns of the
speakers, because they are not shared).

Observe also the use of scare quotes in B's first sentence in (26) (in speech,
this could be marked with special intonation). They serve to distance the speaker
from the expression and thus make it possible for B to talk about the same refer-
ent that A does, *without* undertaking the same expressive commitment. This is, it
would appear, the main function of scare-quotes: to manage expressive commit-
ments.[28]

Other constructions can be used to a similar purpose – Harris (2014) dis-
cusses what he dubs transparent free relatives (TFR). A TFR is the kind of con-
struction used in (27a.), a non-sortal *what* introducing an attributive verb (*call,
think, say* etc.) embedding a noun phrase ("what he dubs transparent free rela-
tives" is also an example, of course). Consider the contrast in (27): in the variant
(a) the speaker uses the term "beergarita" in a transparent way, thereby (implic-
itly) undertaking the commitment to that term being applicable and appropriate.
In the variant (b), however, the speaker – perhaps judging that "beergarita" is an
extremely silly word – goes out of her way not to undertake this expressive com-
mitment and explicitly attributes the use of the term (and therefore the commit-
ment) to John. According to Harris, managing commitments is precisely what
TFRs serve to do.

In those two examples, expressive commitments can appear immediately sa-
lient: in (27), the terms "freedom fighter" and "terrorist" carry heavy loads of as-
sessment and emotion; a speaker using one of them reveals a lot about their po-
litical sympathies. In (28) matters are much less grave, but the name "beergarita"

---

28 I do not elaborate on this thought, but Brandom (1994) does.

stands out (for most users of English at least) as an unfamiliar expression, perhaps meant as a joke. Whether it is to be taken seriously, what exactly it means, and "does anyone actually say that" may be salient questions. However, we can imagine that (27a) is said by one bartender to another, and that "beergarita" is a standard term for a certain type of drink (the ingredients of which are easy to divine) – in such a situation, the expressive commitment to "beergarita" would be no more important than the commitment to "village" in (17). There would be no reason to register it on the conversational scoreboard.

Example (28) shows the converse situation. For a large number of American English speakers "pop" is the standard, unmarked generic term for a sweetened carbonated beverage. Among those speakers, A's way of framing his question would be entirely transparent. However, for many other speakers (the divide is geographical and quite distinct) "soda" is the standard, unmarked term with the same meaning. And so B can, jokingly, bring that up, because for her "pop" is not a default term (and, for unclear reasons, people generally like to tease each other about dialectal differences). A and B's expressive commitments – to the appropriateness of "pop" or "soda" respectively – are relevant because of associations that are not only entirely contextual, but in fact extra-linguistic.

I will call this phenomenon *pragmatic opacity*. An expression is *pragmatically opaque* if it raises the issue of expressive commitment and therefore a commitment to its use must be recorded on the scoreboard. The reasons may be diverse, ranging from lexical to contextual to extra-linguistic; they may have to do with a nuance of meaning of the term, with its unfamiliarity to an interlocutor, with its association with a certain dialect or register etc. Expressions can also be made opaque through the use of scare-quotes, TFRs, or other constructions ("so to speak" etc.). The important point is that an opaque expression brings attention to itself, as it were, or more generally to the way in which a speaker chooses to express herself. The effects of this may also be varied, from manifesting the speaker's attitudes and beliefs ("freedom fighters") to expertise signaling (use of technical jargon) to revealing where the speaker comes from. Sometimes (which will be important later) the opacity is resolved by attributing the expressive commitment to someone else than the speaker (again, constructions like scare-quotes or TFRs are ways of ensuring that).[29]

---

29 Pragmatic opacity should not be confused with referential opacity, but it is in a sense an analogue of it. "The man in the brown hat" and "the man seen at the beach" have the same referent in Quine's classic example, just as "freedom fighters" and "terrorists" do in (26) or "pop" and "soda" in (28) – but it does make a difference which expression is chosen.

Regardless of whom the commitment is attributed to, the attribution can only be resolved if the interpreter, e.g. conversation partner, can infer a reason why the speaker takes the specific expression to be applicable or appropriate[30]. In (26) A's reason for undertaking the commitment to appropriateness of "freedom fighters" is that he or she positively evaluates the actions and goals of the militants; in (27) "beergarita" may be a term of art used by bartenders; in (28) the use of "pop" or "soda" is based on where the speakers happened to grow up. In each case, this inference allows the expression in question to carry additional meaning – that the speaker supports the militants' cause, that beergarita is a thing, that the speaker comes for a certain part of the US. These additional meanings may be more or less conventionalized, but they may also be entirely independent of the semantic content of the expressions in question ("pop" and "soda" are synonymous).

The concept of pragmatic opacity can serve as the key to understanding the nature of expressive meanings. For what is characteristic of expressives as a (fuzzy-bordered) class of lexical items is that they are strongly marked – they are typically taboo swear-words or at least highly colloquial; in any case they are non-neutral and "socially risky", as Lasersohn (2007) puts it, because their use is normally considered inappropriate (even if only mildly).[31] A use of an expressive always constitutes a deviation from a certain standard, a breach of linguistic decorum, and thus is always pragmatically opaque. In other words, expressives inevitably raise the issue of expressive commitment, which needs to be resolved by answering the question of why the speaker takes the expressive to be appropriate. I have suggested that already with regard to example (24) in the previous section ("This theory is bullshit").

My main claim, therefore, is this: expressive meanings result from the attribution of expressive commitments to speakers. They are, therefore, generated pragmatically in the concreteness of the linguistic and extra-linguistic context. By using an expressive a speaker is taking license, as it were, to depart from a

---

30 That is why the issue of expressive commitment is not raised (or trivially resolved) for most of vocabulary – there's no question why someone accepts that a village should be called "a village" or a dog "a dog".

31 A borderline case are words like *totally* or *man* (in their non-descriptive intensifier uses they seem to have purely expressive meanings) – these are not taboo, but they are colloquial, and more importantly, semantically bleached. The latter aspect may account for their pragmatic opacity – a use of a word that contributes no definite semantic content raises the issue of applicability, i.e. expressive commitment. However, as I suggested at the beginning of this essay, it is also possible that this analysis only partially applies to such words because they only partially resemble core examples of expressives such as *damn* or *bastard*.

standard of linguistic decorum; he is now (by default, in non-quotative contexts and barring any perspective shifts, about which see below) committed to this being appropriate. The full meaning of an expressive can only be conveyed as a function of this commitment and the context – as the audience must infer why the speaker takes himself to be warranted in this commitment.[32] (The default interpretation is, for the core examples of expressives, that the speaker is in a heightened emotional state in connection with whatever object the expressive is used to refer to or describe.)

Expressive commitments are thus unlike most of the objects that semanticians concern themselves with – they are not part of a combinatorial construction of truth-conditional meaning, or anything like that. However, they are an important part of discourse dynamics and need to be recorded on a conversational scoreboard, because they can influence the further flow of conversation just as assertions can.

Let me illustrate my account with a simple example of conversational score-keeping for assertoric and expressive commitments. I will use a simple table, in which the interlocutors' (A and B) commitments are registered separately from the common ground, which comprises those commitments upon which both A and B agree, as well as some other material, such as shared presuppositions, standards of precision, mutual beliefs about the extra-linguistic contents, and so on (lets' call the whole set Presupp). I will use double-square brackets for a schematic representation of the propositional contents of assertoric commitments and double curly brackets for expressive commitments (these are not, obviously, formal representations, but can be stand-ins for whatever formal representation one would like to use here). Let A say the following:

A: *Jerry told me some news today.*

A' commitments	Common Ground	B's commitments
[[Jerry told A news]]		
	Presupp…	

---

32 When talking about inferences here, I mean only abstract steps of the construction of meaning in the discourse model; I claim nothing about psychological reality. In practice, the required inferences are probably processed automatically (unless there are some ambiguities as to the object and cause of a speaker's attitude), aided by things such as intonation and gestures.

B responds simply:
A: *Jerry told me some news today*.
B: *Yeah?*

A' commitments	Common Ground	B's commitments
	[[Jerry told A news]] Presupp…	

In this way, B acknowledges A's assertion and admits it into the common ground – both interlocutors accept that Jerry told A some news, so A can continue with B's encouragement.

A: *Jerry told me some news today*.
B: *Yeah?*
A: *That bastard Kaplan got promoted*.

The scoreboard update will be more complex now, so let us divide it into two stages. First, add the assertoric commitment to A's column, just as with his first utterance.

A' commitments	Common Ground	B's commitments
[[Kaplan got promoted]]	[[Jerry told A news]] Presupp…	

But that is not all that A is saying – he has used an expressive, and his commitment to this being appropriate should also be registered:

A's commitments	Common Ground	B's commitments
[[Kaplan got promoted]], {{'bastard' is appropriate}}	[[Jerry told A news]] Presupp…	

We can then imagine that B accepts the news – the assertoric commitment moves to the central column – and focuses on A's strong language:

A: *Jerry told me some news today*.
B: *Yeah?*

A: *That bastard Kaplan got promoted.*
B: *Come on, Kaplan is not that bad. And that's no way to speak of a colleague.*

A's commitments	Common Ground	B's commitments
{{'bastard' is appropriate}}	[[Jerry told A news]], [[Kaplan got promoted]], Presupp…	[[Kaplan is not that bad]], {{'bastard' is not appropriate}}

And so the conversation, and the scoreboard updates, continue. A's expressive commitment becomes a salient part of the conversational record; it's discursive significance is more or less equivalent to a Pottsian proposition **'bad**(Kaplan)' – but instead of stipulating a semantic content of this sort, we can assume that B infers A's negative attitude based on his choice of words.[33] In her response B can both target the inferred attitude and the inappropriate (in her opinion) use of language, but neither of those challenges involves the at-issue propositional content of what A said (that Kaplan was promoted). Expressive meanings are independent.

How does the expressive commitment theory of expressives compare to the multi-dimensional semantics approach? Although in an obviously very different way, my proposed theory can account for the projective behavior of expressives at least as well as the semantic approach.[34] Independence and Nondisplaceability of expressive meanings result not from the fact that they are computed in a separate dimension of the semantic composition and always passed up to the top of the parsetree by special rules (or whatever other implementation one could propose) – but because they do not belong to the semantic composition at all. The expressive meaning is conveyed by a speaker's choice of words, which is an extra-semantic matter (hence Independence), just like the fact that one is speaking a particular dialect of English is an extra-semantic matter (and yet can carry a lot of information about the speaker). And the speaker's choice is the only thing needed to account for Nondisplaceability, because the speaker is by default, in normal circumstances, fully responsible for the words he or she chooses to utter –

---

33 This makes it much easier to accommodate other interpretations – if, for instance, A is clearly just teasing and in truth quite happy that his good friend Kaplan got promoted; "bastard" serves more like an intensifier then, and B can infer this, thus attributing a different meaning to the expressive.

34 It is worth noting that on my account the mechanisms of projection of expressive meanings are very different than the mechanisms responsible for other projective categories such as presuppositions. This is also an advantage, insofar as, intuitively, expressives and presupposition triggers do not have much in common.

regardless of any embedding or displaced context.[35] (And it is always in principle possible to convey the same (at-issue) content without the use of expressives.)

This way of thinking about expressives naturally accounts for the fact that they can only be challenged meta-linguistically, as expressive commitments are, in a concrete sense, meta-linguistic: they are commitments to a certain way of speaking, not (like assertoric commitments) to a propositional content. Descriptive contents can be inferred from the use of expressives (in the discourse update example, *that A holds a negative attitude towards Kaplan* is a descriptive proposition in itself), but they are not what is directly manifested by it. Thus, a meta-linguistic challenge is, in fact, a more direct way of challenging expressives.

One of the problems I pointed out in the semantic theory was that it forces us to turn to *ad hoc* solutions such as type shifts to account for a kind of scope ambiguity characteristic of expressives. By saying *That bastard Kaplan got promoted* speaker A in the update example may be expressing his attitude towards the individual Kaplan or towards the situation as such. It's entirely possible that he would continue the exchange above by saying something like *Yeah, I know he's OK, I'm just upset for being passed over again*. Instead of assuming type shifts, on my pragmatic account we can posit that B's inferences regarding the reasons for A's expressive commitment to *bastard* may be more or less specific. Where the semanticists need to assume one or another meaning for *bastard* (taking an e-type or a t-type expression as argument), and subsequently posit a reanalysis if a different interpretation is forced, we can now simply take the meaning of *bastard* to be underdetermined until the reasons for the speaker's use of strong language are made clear. For the only thing that is fixed with regard to this word is that it warrants attribution of an expressive commitment to the speaker.

A more important problem which I discussed concerned the issue of perspective shifts and non-speaker-oriented interpretations of expressives. The pragmatic account I am suggesting offers a straightforward solution. We have already seen that expressive commitments can be explicitly manipulated: disavowed by using scare-quotes or directly attributed to someone other than the speaker with TFRs. It should be noted that a speech report does not necessarily entail attribution of any relevant expressive commitments to the reported speaker, although it can make it more natural. Consider the following two examples:

---

35 The only regular way of obviating this responsibility is direct quotation – and accordingly, expressives in direct quotes are not taken to express the quoting speaker's attitudes, but the original utterer's.

(30) My 3-year old asked if she could play with my **3D printer** this morning – she called it the 'magic spitting machine', actually.

(31) John said he gave Mary a **beergarita**. Whatever that is.

In (29) the original question of the 3-year old is framed using a term that she apparently hadn't known (and so could not be committed to its appropriateness). This does not make the report infelicitous or deviant – to repeat, the actual speaker is by default responsible for the choice of words, even in reports. However, this default can be circumvented. In (30) the expressive commitment to the applicability of "beergarita" is attributed to John. The felicitous continuation makes clear that the speaker wouldn't even know if the name applied to a given object (a drink, presumably) or not.

Such implicit shifts in expressive commitments are commonplace – and this observation is all that is needed to account for perspective shifts with regard to expressive meanings.[36] Example (3), repeated here again as (31), is very similar to (30). The context makes it clear that the expressive commitment is to be attributed to the subject of the report, as the actual speaker would not use the word "beergarita" or "bastard", respectively. Things are much less clear in (17), repeated as (32), where the expressive is not overtly embedded – which is reflected in the differing interpretations among the respondents in Harris and Potts' original experiment – but it is easy enough to imagine a context in which there are better reasons to attribute the negative expression to the subject than to the speaker.

(32) My father screamed that he would never allow me to marry that **bastard** Webster.

(33) My classmate Sheila said that her history professor gave her a low grade. The **jerk** always favors long papers.

The immediate effect of the use of an expressive is to raise the issue of expressive commitment – to make this choice of words salient and bring the hearer's attention to it. The resolution of this issue may consist in attributing the commitment to the actual speaker (the default option) and inferring a reason why that person is speaking this way – why they take it that strong, perhaps taboo language is warranted in this situation. Or, in suitable circumstances, the issue may be re-

---

36 Still, the speaker default in expressives seems to be stronger than with other subsets of the lexicon – but that, as I argued, results from the inherent opacity of expressives (which in turn is a function of their taboo or low-register status).

solved by assuming that the expressive commitment is to be attributed to someone other than the speaker. This can happen, specifically, when the speech and the manner of expression of another agent is made salient enough in the context. This happens clearly in (31) and potentially in (32) – the point of the reports in both cases is not only what the original speakers said, but also how they said it (the father screamed in a fit of rage, Sheila was extremely upset, etc.).[37]

6 Conclusion

In this discussion I tackled the question of the nature of expressive meaning, insofar as it is revealed in the functioning of expressives. I argued that a multi-dimensional semantic approach, which has become quite influential in recent years, is severely limited and, beyond showing how expressives can be technically integrated in a formal semantic framework, offers no philosophical insight into the aspects of language use that are specific to them. As an alternative, I proposed a pragmatic account which focuses on the most salient feature of expressives as a unique subset of the lexicon and places expressive meanings in a framework of conversational scorekeeping with commitment attribution.

The account I have proposed is indeed multi-dimensional in a stronger and more specific sense than the Pottsian logic of conventional implicature. The latter assimilates expressive meanings to descriptive ones, only ensuring that they remain separated in the semantic composition. On my view, expressive meanings turn out to be of a very different kind than descriptive meanings – they are constituted at a different level, as it were. It is the level of commitment attribution, not of truth-value computation; it concerns more directly what people do with their words than what those words denote.

In this sense, the theory of expressive meaning I propose here is also a more direct implementation of the idea of *meaning as use*. The meanings of expressives are constituted directly in their use on this account: it is by choosing to utter those very words that a speaker expresses her attitudes and emotions. If we look at expressive meanings this way, it is possible to extend this line of thought beyond the relatively narrow class of words such as *bastard* and *damn*. Slurs (racial, ethnic, homophobic etc. insults) illustrate a potential application: they do not carry any specific descriptive contents, but in choosing a derogatory word for a given

---

37 See Hess (2018) for an elaboration of this aspect of the pragmatic account and a discussion of the factors potentially contributing to perspective shifts of expressive meanings.

group of people instead of a neutral, non-offensive term, a speaker makes her attitudes towards the target group manifest by undertaking an expressive commitment to the slur's being appropriate. Other categories may also be included, even non-lexical ones like politeness – expressivity is, after all, a pervasive feature of language use. Expressives in the narrow sense are simply specialized tools to bring expressivity and subjectivity of speakers to the foreground.

References

Amaral, Patricia/Roberts, Craige/Smith, E. Allyn (2007): "The Logic of Conventional Implicatures." In: *Linguistics and Philosophy* 30, pp. 707–749.

Bach, Kent (1999): "The Myth of Conventional Implicature." In: *Linguistics and Philosophy* 22, pp. 327–366.

Barker, Chris/Bernardi, Raffaella/Chung-chieh Shan (2010): "Principles of Interdimensional Meaning Interaction." In: *Proceedings of SALT* 20, pp.109–127.

Brandom, Robert (1994): *Making it Explicit*, Cambridge, MA: Harvard University Press.

Farkas, Donka/Bruce, Kim (2010): "On Reacting to Assertions and Polar Questions." In: *Journal of Semantics* 27, pp. 81–118.

Grice, Herbert Paul (1975): "Logic and Conversation: In: Peter Cole/Jerry L. Morgan (Eds.): *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press, pp. 41–58,

Gutzmann, Daniel (2015*): Use-Conditional Meaning: Studies in Multidimensional Semantics*. Oxford: Oxford University Press.

Harris, Jesse/Potts, Christopher (2009): "Perspective-shifting with Appositives and Expressives." In: *Linguistics and Philosophy* 32, pp. 523–552.

Harris, Jesse (2014): "Signaling Non-Speaker Commitment in Transparent Free Relatives: A Paired Speaker-Hearer Judgment Study." In: Verena Rieser/Philippe Muller (Eds.): "*The Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue."* Edinburgh, pp. 71–79. Available at: http://www.macs.hw.ac.uk/InteractionLab/Semdial/semdial14.pdf

Hess, Leopold (2018): "Perspectival Expressives." In: *Journal of Pragmatics* 129, pp. 13–33.

Horn, Lawrence (1985): "Metalinguistic Negation and Pragmatic Ambiguity." In: *Language* 61, pp. 121–174.

Kaplan, David (1999): *The Meaning of Ouch and Oops*. Howison Lecture in Philosophy delivered at UC Berkeley. Available at: http://eecoppock.info/PragmaticsSoSe2012/kaplan.pdf

Lasersohn, Peter (2005): "Context Dependence, Disagreement, and Predicates of Personal Taste." In: *Linguistics and Philosophy* 28, pp. 643–686.

Lasersohn, Peter (2007): Expressives, Perspective and Presupposition." In: *Theoretical Linguistics* 33, pp. 223–230.

Lewis, David (1979): "Scorekeeping in a Language Game." In: *Journal of Philosophical Logic* 8, pp. 339–359.

Potts, Christopher (2005): *The Logic of Conventional Implicatures*. Oxford: Oxford University Press.

Potts, Christopher (2007): "The Expressive Dimension." In: *Theoretical Linguistics* 33, pp. 165–198.

Schlenker, Philippe (2007): "Expressive Presuppositions." In: *Theoretical Linguistics* 33, pp. 237–245.

Pedro Abreu
Temperate Semantic Conventionalism

Abstract: I return to Davidson's "anti-conventionalism" papers to assess his famous arguments against the sufficiency and necessity of conventions for successful linguistic communication. Davidson goes beyond the common contention that the basic conventional layer of meaning, one that is secured by interlocutors' shared competence in their common language, must often be supplemented in rich and inventive ways. First, he maintains that linguistic understanding is never exclusively a matter of mere decoding, but *always* an interpretative task that demands constant additional attention to the indeterminately various cues and clues available. More radically still, Davidson denies that linguistic conventions are even needed. In particular, he argues against the fairly consensual thesis there is some essential element of conventionality in literal meaning. This still represents a very distinctive contribution to the persistent and tumultuous discussion over the relative natures and limits of semantics and pragmatics.

I maintain that Davidson is only partially right in his claims. I agree with him about the general insufficiency of conventions for linguistic communication. I develop an argument supporting the thesis that genuine pursuit of linguistic understanding can never take the form of uncritical conformity to a fixed norm. I am also convinced that Davidson is right about the occasional dispensability of conventions. Often enough, as Davidson's examples show, literal meanings are improvised on the go – that is, interlocutors manage to coordinate on the meaning of some exchanged expression without the benefit of a shared convention governing that use. I reject, however, general non-necessity. I consider in some detail Davidson's argument from radical interpretation and conclude that it fails.

Keywords: semantic conventionalism, semantic anti-conventionalism, Donald Davidson, literal meaning, malapropisms, radical interpretation

1 Introduction

There is a tempting view of linguistic practices and interpretation that Davidson identifies and denounces as inadequate in a number of texts from the eighties and nineties. Let us call it "the rigid conception" of linguistic communication. According to it, "in learning a language, a person acquires the ability to operate in accord with a precise and specifiable set of syntactic and semantic rules; verbal

communication depends on speaker and bearer sharing such an ability, and it requires no more than this" (Davidson 1994, p. 110).

This view is endorsed in different forms, to different extents, by different authors.[1] The crucial commonality is the placing of some notion of *regularity*, *convention* or *rule* at the core of our linguistic competence, and the reduction of interpretation to the exercise of a very specific, and potentially automatic, ability – that of calculating the standard meanings of uttered expressions in accordance with some system or code commonly known by all participants. Such a straightforward method would release interpreters – at least in some cases – from the need for any additional inquiry into the speaker's mind and background. Davidson argues, against this view, that the notions in question are not so important or central to successful linguistic interchanges.

In Section 2, I circumscribe the discussion to Lewis' notion of *convention*. I develop in some detail Lewis' convincing analysis of these social structures, and elaborate on how they can be used to support and facilitate communication and interpretation. In Section 3, I turn to a second notion, *first meaning*. Davidson motivates it first as a key piece in his account of situations where the interchange of literal meanings among speakers is successful despite the fact of words not being used to carry their conventional or standard meaning. Equipped with these notions, we'll finally be in an adequate position to present, in Section 4, Davidson's specific form of anti-conventionalism. He does not dispute the pervasiveness and practical convenience of conventions of meaning in verbal communication. What he disputes is that there is a more intimate relation than that. What he disputes is the thesis that language and linguistic practices are essentially conventional.

In Section 5, I finally turn to the arguments behind Davidson's position. Davidson invokes malapropisms, and other related occurrences of semantic innovation, to make a solid case, illustrated with abundant examples, in favor of occasional and local insufficiency and non-necessity of conventions. How far can these results be extended? In Section 5.1, I follow an argument in favor of general insufficiency. The argument, latent in Davidson's framing of the discussion, is first picked up by Pietroski, and then reformulated by Lepore and Ludwig. I introduce some readjustments and conclude that the argument successfully supports the general claim. In Section 5.2, I consider the argument from radical interpretation in favor of general non-necessity. This is the clearest argument

---

1 Aristotle and Locke are often referred as classical precursors of the view. Very recently, Ernest Lepore and Matthew Stone, offer an especially pure formulation of the doctrine: *Imagination and Convention* (2015).

available. It is, at least, strongly suggested by Davidson, and often identified in the relevant literature. I contend that the argument fails.

In defense of an alternative conception, he sets himself to show that conventions are neither sufficient nor necessary for linguistic communication. I present and inspect several (more or less straightforwardly) Davidsonian arguments to that effect. I conclude that there is a good case for the general insufficiency of conventions in interpretation, but that the arguments for non-necessity fall short of demonstrating a strong and general dispensability of conventions.

2 Convention

There is a discernible emphasis on *convention*, even if Davidson often shifts from that notion to others closely related – that is, there is also talk of *regularities*, *rules*, *norms*, and maybe more. I will focus exclusively on convention as it appears to be the ultimate target of Davidson's arguments and positions.

Davidson, along with most others in this discussion, refers to David Lewis' analysis of the notion (Lewis 1969 and 1975). According to Lewis, conventions arise as rational solutions to coordination problems. Coordination problems are problems involving more than one agent, where there is a predominant coincidence of interest among the participants, and where the success of each agent in reaching some desired outcome is dependent upon her ability to concert her course of action with those of the other participants. Participants must form their strategies based on their expectations concerning others. No individual move can warrant a favorable result in such situations; only coordinated actions can pull it off. Lewis provides several simple and illuminating examples of coordination problems such as *meeting someone*, *resuming an interrupted phone conversation*, *rowing together*, *driving on two-lane roads*, and more (Lewis 1969, p. 5 ff.).

For a coordination problem to call for a conventionalized solution, another decisive feature must be in place: the problem must allow for various solutions equally convenient for those involved. This corresponds to the element of arbitrariness in conventions. If there were good enough independent reasons for agents to converge on a particular solution, a convention would not be required for the same effect. Consider one of Lewis's examples, the coordination problem of how to share a two-lane road. What all drivers wish to avoid is that "some drive in the left lane and some in the right" (ibid. 6), because then "everyone is in danger of collision" (ibid.). But for each driver "[i]t matters little ... whether he drives in the left or the right lane, provided the others do likewise" (ibid.). There being

no independent reasons determining a concerted response, this is precisely the kind of case that justifies a convention.

For a convention to arise, it must first happen that some particular coordination strategy becomes collectively focused. This can happen in a number of ways, from explicit agreement to some accidental precedent, or natural saliency. The rest follows rationally. Agents that start with adequate expectations about each other and who share a common interest choose their actions in accordance and manage to coordinate and attain common success. Good results reinforce initial mutual expectations, new successes grow more likely each time, and participants' behavior becomes more and more regular. At some point, a *convention,* in the Lewisian sense, will come to be in force.

Here is a late version of Lewis' analysis, with some trimming:

> A regularity *R*, in action or in action and belief, is a *convention* in a population *P* if and only if, within P, the following six conditions hold. (Or at least they almost hold. A few exceptions to the "everyone"s can be tolerated.)
> 1.   Everyone conforms to *R*.
> 2.   Everyone believes that the others conform to *R*.
> 3.   This belief that the others conform to *R* gives everyone a good and decisive reason to conform to *R* himself. (...)
> 4.   There is a general preference for general conformity to *R* rather than slightly-less-than-general conformity. (...)
> 5.   *R* is not the only possible regularity meeting the last two conditions. There is at least one alternative *R′* such that the belief that the others conformed to *R′* would give everyone a good and decisive practical or epistemic reason to conform to *R′* likewise; (...)
> 6.   Finally, the various facts listed in conditions (1) to (5) are matters of *common* (or *mutual*) *knowledge*: they are known to everyone, it is known to everyone that they are known to everyone, and so on.
>
> <div align="right">(Lewis 1975, pp. 164–165)</div>

We can see how it fits perfectly well our previous example. Take, for instance, continental Europe. Every driver drives in the right lane, and everyone expects all other drivers to do the same. The belief that all other drivers will be driving in the right lane gives each driver a good and decisive reason to drive in the right lane. No one would profit from exceptions to this regularity – that is, if most people already drive in the right lane, everyone prefers that all drivers do the same. It would do just as well if all drivers were instead to drive in the left lane. The original choice between left or right is arbitrary as there are no (evident) reasons

to prefer one side or the other. Lastly, all this is common knowledge[2] to all involved: all drivers know all the facts stated above, and know that all drivers know them, and know that all drivers know that all drivers know them, *and so on*. This last condition is meant to *stabilize* the convention. Remember that these are situations where agents must choose their actions based on their expectations about the other agents. Common knowledge of all these facts allows each agent to replicate the others' reasonings and, in general, to grasp the rationality and goodness of the whole arrangement, hence confirming the agent in the persuasion that driving in the right lane is both the thing to be expected and the right thing to do.

How are conventions thus defined to play a role in linguistic communication? Lewis himself applies the notion as the basis of his own foundational semantics. He affirms a conventional relation of linguistic expressions to their meanings. According to his proposal in "Languages and Language", what relates a speaker to her particular language, or idiolect, and her words and sentences to their particular meanings, is a convention of *truthfulness and trust* in that language. This convention is defined as in the above general scheme by replacing "R" with "a regularity of truthfulness and trust in £", where £ is a language, characterized as complex function from expressions to their meanings or truth-conditions, and the conformity to the regularity in question consists of "members of P frequently speak[ing] ... sentences of £ to one another", with speakers trying to utter only sentences they believe true in £, and speakers responding to such utterances "by coming to share th[e relevant] belief ... and adjusting [their] other beliefs accordingly" (Lewis 1975, p. 167).

For present purposes, we can make things simpler and more perceptible by letting go of *trust and truthfulness* and invoking instead *a convention of meaning in £*, where the regularity in question would simply be that of using expressions of £ in accordance with their meaning in £ — that is, if someone utters some linguistic expression *e,* that person (expression) means by it what *e* means in £; if someone hears it uttered, that person responds by coming to take the speaker to have (expression) meant by it what *e* means in £.

However, what we shall be focusing on are not so much the global commitments attached to entire languages, but subsidiary conventions, in some form included therein, that govern the use of each particular expression of the language. Take, for instance, "Snow is white". Following the previous line, we should say that

---

2 Lewis settles for *potential* common knowledge, that is, knowledge "available if one bothered to think hard enough" (Lewis 1975, p. 165). I agree with Daniel Nolan when he notes that condition (2) seems somewhat "redundant, given that all the conditions must be common knowledge in the population P" (Nolan 2005, p. 161).

a regularity *R*, in action or in action and belief, is a convention, in a population P, of *using "Snow is white" to (expression) mean that snow is white* if and only if,

1. The members of P conform to the regularity of taking "Snow is white" to (expression) mean that snow is white. That is, if someone utters "Snow is white" that person (expression) means that snow is white; if someone hears it uttered, that hearer responds by taking the speaker to have (expression) meant that snow is white.
2. Everyone believes that the others conform to the regularity of taking "Snow is white" to (expression) mean that snow is white.
3. This belief that the others conform to *the regularity of taking "Snow is white" to (expression) mean that snow is white* gives everyone a good and decisive reason to conform to it.
4. There is a general preference for general conformity to the regularity of taking "Snow is white" to (expression) mean that snow is white.
5. The regularity of taking "Snow is white" to (expression) mean that snow is white has alternatives.
6. All these facts are common knowledge in P.

With this notion of convention in place and these examples, we are now in a better position to consider Davidson's claims about convention's dispensability and insufficiency for linguistic communication. First, however, we must clarify the notions of linguistic communication and interpretative success at stake in his arguments and claims, as well as the central notion of meaning involved.

3 First meaning

When the question is the role of convention in linguistic communication, we must be clear about what defines success in linguistic communication. That definition must be plausible while, at the same time, it must not commit us straightaway to a position for or against Conventionalism. Davidson's notion of "*first meaning*" (Davidson 1986, p. 91) plays the central role in it. Let us uncover its features.

To begin with, the notion we're after is an intention-based notion. Davidson is clear about his conviction that the crucial thing in meaning and communication is being understood as one intends to be understood – coordination is as much at the root of Davidson's view as it was at the root of Lewis'. In his words, the "intention to be taken to mean what one wants to be taken to mean is ... the

only aim that is common to all verbal behavior" (Davidson 1994, p. 120).[3] All successful instances of communication involve the satisfaction of some communicative intention and all conveyed (non-natural) meaning must be specified there, in its correspondent intention.

At the same time, first meaning is intended to be full-fledgedly linguistic. First meanings are articulated by means of expressions integral to a complex system, a language,[4] that allows for the distinctive properties of compositionality, systematicity, and creativity that Davidson himself had such a leading role in calling our attention to (Davidson 1965, 1967). First meanings emerge at a crucial intersection between mind and language. They first link interpersonal understanding to a verbal code, dramatically improving the first and infusing the second with actual content.

The notion of first meaning also brings with it the possibility of a more apt and organized conceptual landscape. Davidson makes clear that we must preserve the traditional distinction "between what a speaker, on a given occasion, means, and what his words mean" (1986, p. 91). However, this distinction should not be secured at the expense of two inadequate conflations, both of intended meaning and non-literal meaning and, especially, of literal meaning and conventional meaning. As Davidson notes, phenomena such as *malapropisms* require us to acknowledge both that literal meaning can happen outside conventions, social institutions and standards, and in direct dependence upon the speaker's communicative intentions.

Linguistic expressions are frequently misused, or used in innovative ways, without the loss of communication and understanding – malapropisms are but one notorious example of that. Speakers often employ words intending them to mean something different from their *standard meanings* and "get away with it" (Davidson 1986, p. 98), i.e. are understood by their audiences as meaning what they actually intended to mean. Mrs. Malaprop produces an utterance of "a nice derangement of epitaphs" and is successfully interpreted as meaning what she intends to mean, i.e. *a nice arrangement of epithets* (Davidson 1986, pp. 103–104). Archie Bunker uses "monogamy" and is successfully interpreted as meaning what he intended to mean, i.e. *monotony* (Davidson 1986, p. 90). Goodman Ace (Davidson 1986, p. 89) explores similar replacements – using, for instance, "granite" instead of "granted" and "baffle" instead of "battle" – only he does so intentionally and much more densely.

---

3 See also Davidson (1993, p. 171, and 1986, pp. 92–93 and 98–99).

4 Taken not in the same sense that Davidson has in mind when he issues his anti-conventionalism by denying the existence of languages (cf. Davidson 1986, p. 107).

With *first meaning,* Davidson is forging a notion that allows for the distinction between *literal meaning* and *speaker meaning*, while accommodating the fact that the former might not be standard or conventional – *epithets*, *monotony*, *granted* and *battle* are meant literally in the cases above – but conferred upon the words by means of a communicative intention of the speaker that is adequately recognized by the audience. This even permits us to make sense of cases where speakers have their words understood in non-conventional and non-standard ways and still manage to add to that some extra layer of meaning. Lepore and Ludwig (Lepore and Ludwig 2005, pp. 265–266) elaborate an example. Someone could comment upon Mrs. Malaprop's words, reproducing her own misuses, by saying "And that's a nice derangement of words". In the imagined context, the utterance's *literal* or *first* meaning would be "And that's a nice arrangement of words", but the speaker intends it to be taken as an ironical remark.

The first meaning of an utterance will often correspond to the expression's standard meaning. Often, the first meaning will "come first in the order of interpretation" (Davidson 1986, pp. 91–92). Ultimately, however, it is the speaker's intention that specifies it. First meanings are determined by what the speaker intends her[5] words to mean, i.e. by what Davidson refers to as the speaker's "semantic intention" (Davidson 1993, pp. 170–171),[6] but are only confirmed, so to speak, if the audience manages to recognize and adequately handle that intention.[7]

We are now equipped with the notion of linguistic meaning and the model of successful linguistic communication that matter to our discussion. *First meaning* corresponds to a notion of the *intended, potentially non-conventional and non-standard, literal meaning of an utterance*. Interlocutors manage to communicate linguistically at the basic level here in question when they manage to converge on the first meaning of their utterances. For the purposes of the present discussion, I will accept Davidson's position on the primacy of the notion of *first meaning* and on what constitutes successful linguistic communication without further discussion[8], and assess his anti-conventionalist arguments in those terms.

---

5 Whenever possible, to reduce ambiguity, I will reserve male pronouns for *interpreters* and female pronouns for *speakers*.

6 Cf. Davidson (1986, p. 92; 1993, pp. 172–3).

7 See Davidson (1986, p. 98 ff.). Below I'll return to the role of the audience in the determination of first meanings.

8 The actual adequacy of these positions is directly challenged by various authors. See, for instance, Dummett (1986), Green (2001), Reimer (2004) and Camp (2016).

4 Davidson's Anti-Conventionalism

Davidson's position on meaning conventions is complex; he expresses different forms of rejection to different conceptions of an alleged conventional nature of linguistic practices. First of all, he never rejects that there actually are meaning conventions, nor that they are pervasive and quite useful in real instances of communication (Davidson 1994, p. 110; 1982, p. 278). What Davidson rejects is, on the one hand, the *necessity* of such conventions for communicative purposes and, on the other, its *sufficiency* (1994, p. 110; 1982, pp. 278–279).

Second, we must also pay attention to the different types of conventions to which he refers. We must distinguish at least two types of regularities considered by Davidson. The first type of regularity considered determines a strong form of sharing. It consists of "speaker and hearer mean[ing] the same thing by uttering the same sentences" (Davidson 1982, p. 276). Davidson defends that "such conformity, while perhaps fairly common, is not necessary to communication" (Davidson 1982, p. 276). He immediately points to the obvious counterexamples: "[e]ach speaker may speak his different language, and this will not hinder communication as long as each hearer understands the one who speaks" (Davidson 1982, p. 276). Examples include cases of speakers of officially different languages and cases of speakers that consistently pick different words and formulations within the same official language or dialect.

Davidson moves his focus to a second type of regularity. It determines a weaker form of sharing and consists of "speaker and hearer … assign[ing] the same meaning to the speaker's words" (Davidson 1982, p. 277). This time, it does not matter which expressions each interlocutor picks in her utterances, as long as both speaker and hearer are ready to interpret them in the same way. This is the type of regularity we will be focusing on. Not by accident, that is the type of regularity that is involved in meaning conventions of the kind exemplified above for "Snow is white."

Two further ideas are important to bear in mind. First, as Davidson notes: "Regularity in this context must mean regularity over time, not mere agreement at a moment" (Davidson 1982, p. 278). This is necessary to secure a distinction between full participation in a meaning convention and mere momentary convergence upon the first meaning of some utterance. Second, whether or not the regularities in question correspond to any official language or constitute some standard in a wide enough community is not a particularly discernible worry in Davidson's approach. He centers his argumentation on the general case, conventional meanings, of which *official language conventions* and *large community language conventions* would constitute the most typical examples.

Davidson states his anti-conventionalism also in a different guise. In "A Nice Derangement of Epitaphs", he famously declares that "there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed" (Davidson 1986, p. 107). Naturally, Davidson has no intention of refusing the existence of language, or even languages, *tout court*. What he is refusing is only a very specific conception of language, one that is fit to play the central role in what I have above called the rigid conception of linguistic communication. According to this conception, meanings involved in instances of successful verbal communication must be *systematic*, *shared* and *governed by learned conventions or regularities* (Davidson 1986, p. 93). What Davidson is refusing is the existence of languages as something suited to determine meanings possessing the three characteristics just listed – that is, languages as stable systems of some sort, knowledge of which would be generally shared among the competent speakers of the same linguistic community, and exclusively relied upon for getting to the literal meaning of any utterance produced among them.

Davidson defends that we should let go of the third feature. Meanings, in his alternative conception of linguistic communication, must still be systematic and shared. Only a rich and sufficiently articulated system can offer the expressiveness characteristic of verbal communication. Furthermore, speaker and interpreter still need to converge on the first meanings of the words used during their interchange if communication is to count as being effective. What Davidson questions is the need for such a convergence to be secured beforehand by means of both interlocutors bringing to the conversation an acquired, common, and standing knowledge of the conventions and regularities of the language. Davidson is convinced that the frequent cases of successful communication undisturbed by semantic innovations – which he illustrates with his examples of malapropisms – force us to reconsider and ultimately eliminate the requirement.

Davidson formulates his revision of the rigid conception with the help of the notions of *prior theory* and *passing theory* (Davidson 1986, pp. 100–101). The theories in question would be descriptive meaning theories specifying the meaning for every expression in the linguistic system of reference – truth theories if we follow Davidson's specific approach to the task. Prior theories model the competence or dispositions to linguistic behavior that each participant brings to the conversation. The interpreter's prior theory should characterize his dispositions to interpret his interlocutor prior to actual conversation, and the speaker's prior theory should characterize how she expects the interpreter to understand her words. Prior theories, Davidson maintains, may be discrepant without any loss of communication. Passing theories model verbal dispositions during the interchange, after any relevant news being gathered and processed. The interpreter's

passing theory characterizes how he actually ends up interpreting the speaker's words, while the speaker's passing theory characterizes how she intends him to interpret them. It is passing theories that need to coincide if communication is to be successful. And that is enough to ensure that speakers' words are taken to mean what they intend them to be taken to mean.

According to Davidson's revised view, a language is at most a fleeting thing, fixed only for the instant of communication, and not a stable system shared among all speakers and accounting for their linguistic competence, something to be "learned, mastered or born with" (Davidson 1986, p. 107). What speaker-interpreters carry with them from situation to situation, allowing them to reach a good enough coordination each time, cannot be identified with either prior (Davidson 1986, p. 103) or passing theories (Davidson 1986, p. 102). It is not, in fact, something that could fairly be described as knowledge of a language, whether one imagines it encapsulated in a module or the cognitive possession of the whole mind and person. Even if rooted and dependent upon a general and structural competence comprising "a basic framework of categories and rules, a sense of the way English (or any) grammars may be constructed, plus a skeleton list of interpreted words for fitting into the basic framework" (Davidson 1986, p. 104), speaker-interpreters' performances reveal more than the simple employment of any form of standing knowledge. They reveal the continuous exercise of a dynamic, complex, and high-order capacity, the person's general intelligence, able to serve in communication as well as in other forms of theory building and mundane navigation (Davidson 1986, p. 107). Mindreading is, of course, central in the affair, as the point is always either to get to our interlocutor's communicative intentions or to render one's own adequately accessible. A rich and flexible notion of rationality mediates the encounter and guides the process both ways. To sum it up, we can say that what speaker-interpreters' performances exhibit is the constant practice of rationality and rationalization, one that is supported in all their accumulated knowledge and resources, unrestricted to any special domain – that is, only part of it is strictly linguistic.

The picture just presented constitutes an enlarged rendition of the same anti-conventionalist position described above. In allowing prior theories to diverge, Davidson sustains the insufficiency and non-necessity of conventions and previously learned regularities for communication. The conventions that are in place, where a speaker's and an interpreter's prior theories intersect, are not enough to ensure communication. Prior theories must be revised or confirmed into passing theories. Speakers achieve this not with the help of further conventions and strict strategies, but by making free and ingenious use of their broad rationality in exploring varying and non-chartable provisions of miscellaneous knowledge.

5 Arguing against Conventionalism

It is time now to turn to Davidson's arguments against conventionalism. In support of his claims that conventions are neither necessary nor sufficient for linguistic communication, Davidson invokes the fact that, sometimes, speakers must figure out the meaning of their interlocutors' words without relying on previous knowledge of any prevailing convention. Speakers can do it, and often have no other option (Davidson 1982, p. 278; 1986, pp. 89–90; 1994, p. 115ff.). Malapropisms are taken as central cases in this argument. The examples elaborated above already go to show that, at least occasionally and locally, speakers manage to communicate and understand each other by exploring other means besides shared knowledge of the relevant meaning conventions. There being no linguistic convention of using "baffle" to mean *battle,* knowledge of linguistic conventions is flagrantly *insufficient* to explain how it can happen that on certain particular occasions speaker and hearer manage to converge in this interpretation. From the opposite angle, even without the knowledge of the convention of using "arrangement" to mean *arrangement*, as presumably was the case with Mrs. Malaprop, a speaker can still be understood by her interpreter to mean it – thus showing how conventions can also be *unnecessary*. The important question now is how far these conclusions can be extrapolated. Is there a valid inference from occasional to any general form of insufficiency and dispensability of meaning conventions in communication?

5.1 Arguing against Sufficiency

Are conventions and knowledge of conventions enough to explain and allow for linguistic communication in its familiar form? The most flagrant counterexamples to that hypothesis involve expressions for which there is a standard meaning and a widespread convention governing their use in the community, but which the speaker manages to employ successfully in some alternative way with some non-conventional and non-standard meaning. More generally still, we can also say that each case of successful verbal communication involving the employment of expressions in non-conventional and non-standard ways is a case of insufficiency of conventions, standards, and their correspondent knowledge for communication. Davidson has this argument in mind when he illustrates the existence of such cases of non-conventional successful communication with the

examples provided above. Apart from this, however, he offers no other clear argument against insufficiency, and actually appears much more interested in the second aspect of the problem: necessity.

As things stand, the insufficiency of conventions and standards is secured only at a very local and occasional level. Some extraordinary occasions require speakers to coordinate beyond what is previously fixed and shared among them. But it would seem that, outside of those rare occasions, learned conventions and standards still suffice for communication. We can easily agree that the occasions might not be so rare after all, and that they encompass more than malapropistic uses. There are many other cases of improvisation and innovation: new names, new words, new idioms, new uses. Nonetheless, if insufficiency and the opportunity or need to invest more in communication than one's knowledge of fixed standards and conventions were keyed only to such special moments of innovation, they would still be confined and of little expression in the whole picture of language use.

Pietroski, for one, in a commentary sympathetic to Davidson's position, expands on the topic and suggests an argument supporting the generalization of insufficiency. It appears in a single paragraph that is worth quoting at length.

> Once the distinction between prior and passing theories has been drawn, I take it that the conceptual distinction matters even if the deliverances of prior and passing theories are the same. Not modifying a prior theory is, on my reading of Davidson, just as much an interpretive decision as making a modification; though in the former case, the 'decision' will typically not be associated with any conscious processing or feelings of 'dissonance.' Successful communication is always a matter of converging *passing* theories; and general intelligence is always implicated here, if only by giving 'tacit approval' to the deliverances of prior theories. Moreover, even if such tacit approval is often granted, one cannot speak of rules *governing* passing theories. For the 'rules' could always be overridden in cases of the Malaprop/Donnellan sort; and one cannot capture the extent of these cases formally or in advance.
>
> (Pietroski 1994, p. 105)

Pietroski points out that the simple drawing of the conceptual distinction between prior and passing theories creates room for the possibility of a permanent actualization of linguistic dispositions taking place when speakers engage in verbal communication, whether or not changes are actually consummated in the process. He claims that an "interpretive decision", explicit or tacit, is in any case called for, be it to confirm or modify the prior theory, and that this can be so even if the interpreter is unaware of it, phenomenologically speaking. Lastly, he stresses the fact that *the last word* does not belong to the conventions or rules of language, which again supports the idea that there is always interpretation as

well as central and personal control of the interpreter over how to interpret any utterance.

Pietroski is spare in details. Lepore and Ludwig, although close enough to the main line of the argument, expand the discussion in two new directions. They stress the epistemological tone of the argument by bringing *justification* to the fore, and they raise new questions about how to understand *sufficiency/insufficiency of conventions*. The critique of both developments will allow me to arrive to a clearer understanding and a stronger version of the argument.

Lepore and Ludwig find the following argument in Pietroski's passage: since "the possibility a speaker has not used his words in conformity with public norms is ever-present" (Lepore and Ludwig 2005, p. 270), interpreters are called to *justify* their option in either way, that is, whether they decide to go with the conventional interpretation or not. "But since [this] justification will invoke more than knowledge of conventions for the use of words, and even that the speaker is a member of the appropriate linguistic community, it follows that knowledge of conventional meaning is never sufficient for interpretive success" (Lepore and Ludwig 2005, p. 271).

They acknowledge some truth and cogency in this line of reasoning but, at the same time, they note that "it also seems clear that we routinely and successfully interpret others on the basis of taking them to mean what their words mean according to public norms" (Lepore and Ludwig 2005, p. 271). They conclude that, ultimately, it all "boils down to what we intend by saying that knowledge of conventional meanings is sufficient for interpretive success" (Lepore and Ludwig 2005, p. 271). If what it meant is

i.   "that sometimes, even often, we are not called upon to revise our view that the speaker speaks with the majority" (Lepore and Ludwig 2005, p. 271).

then the argument does not secure generalized insufficiency of conventions. If, on the other hand, what is meant is

ii.   "that knowledge of conventional meanings all by itself sometimes suffices for interpreting another as speaking in accord with public norms" (Lepore and Ludwig 2005, p. 271).

then we can take the argument to show that conventions are indeed always insufficient for communication. This is Lepore and Ludwig's take on Pietroski's argument. As anticipated above, I would like to press a bit further on the two new developments.

First, I want to challenge the effective availability of alternative interpretations of the hypothesis here at stake – i.e., general insufficiency – that cor-

respond to the alternative interpretations of its counterhypothesis – i.e., general sufficiency – put forward as *i*. and *ii*. in the previous paragraph. Starting with *i*., it is so flagrantly the case that speakers often conform to conventions and must be interpreted in accordance with them that to question this would be pointless. This form of occasional sufficiency appears beyond dispute, and no interesting and plausible hypothesis of general insufficiency could be offered against it. The relevant question cannot be whether or not *i*. Reading *ii*., on the other hand, is too open and vague – as Lepore and Ludwig themselves note, insufficiency in this sense would follow from an observation as general and anodyne as the fact "that human beings are in general fallible" (Lepore and Ludwig 2005, p. 271).

I don't think we're so clueless about what's really the matter here. The pertinent question is whether the rigid conception of linguistic communication is still an acceptable account of what takes place at those occasions where meanings are meant and ascribed in accordance with the prevailing conventions, or whether that conception is generally undermined. The pertinent question is whether interpreters are at any time actually relieved of any supplementary work of deeper and complex interpretation, and permitted to rely exclusively on their knowledge of the conventions. Even at those instances where there is no deviation from the communal norm and no revision of the standard or conventional meaning is ultimately called for, interpretation still demands the practice of rationality and rationalization as briefly introduced in the previous section – this is the Davidsonian thesis that genuinely deserves attention and scrutiny.

Second, I want to recommend a change in the argument. I believe the appeal to justification, central in the argument recovered by Lepore and Ludwig, is superfluous and misleading. We can make it sharper if we let it go. It is not only that knowledge of conventions is insufficient to sustain a conventional interpretation of an utterance *in a justified way*; it is more simple and more serious than that. It is that knowledge of conventions is insufficient to sustain a conventional interpretation of an utterance even without the qualification.

As Lepore and Ludwig articulate the *justification* problem, they make it sound as if justifying one's interpretive decision or not is somewhat optional. They even explain, in a footnote, that the argument is conditioned upon a specific epistemological position that liberally affirms the need to justify every interpretation. They point out that this view can be challenged: "it might be maintained that our beliefs about what others in our community mean by their words and actions are justified by default: unless circumstances depart from the norm, in some way that we should notice, the beliefs we have automatically are justified without appeal to anything" (Lepore and Ludwig 2005, p. 271, fn. 220) In

this contrary conception, only sporadically would justifications be called for, and ascriptions would need to "be actively justified only when circumstances depart in certain specific ways from the norm" (Lepore and Ludwig 2005, p. 271). I believe we can form a tighter argument for the general insufficiency of conventions, and one that is not dependent on disputable epistemological stances, by just letting go of the idea of justification.

The revised argument I propose stems from the same crucial observation that there is an ever-present possibility that the "speaker has not used his words in conformity with public norms" (Lepore and Ludwig 2005, p. 270). Sometimes, interpreters detect that this is indeed the case – straightforwardly, this gives us occasional insufficiency. Now, the crucial inferential step towards general insufficiency is supported by the consideration that any occasional detection of non-conventional uses already requires some prior form of attention to that possibility. That is, it cannot be that the detection of deviation triggers deeper interpretation because deeper interpretation is already needed for detecting deviation. The interpreter would not detect any departure from the conventional course if he were not both aware of the possibility of deviant uses and on the lookout for them, however unconsciously and effortlessly. He would simply continue ascribing standard meanings, and suffer, unwittingly, the consequent losses in understanding. This is not what happens – interpreters do detect anomalies and reinterpret in accordance. Furthermore, I must stress that *every* utterance is a chance for deviation. Because of that, interpreters are required to be continuously on the lookout. Virtually never are they allowed to rest exclusively on their knowledge of the community's norms and conventions. What emerges here, I conclude, is the generalized insufficiency of conventions and the correlative need for interpreters to engage in a continuous exercise of rationality and rationalization.

5.2 Arguing against Necessity

Can speakers do without conventions in their linguistic practices? Davidson's examples of malapropisms clearly show that sometimes speakers manage to understand each other by means other than the exploration of meaning conventions. Once more, the important question that follows is how far can we extend this conclusion? Is this merely an exceptional phenomenon, or is it possible to engage in verbal communication while completely dispensing with the use of conventions?

The examples explored hint at the weaker conclusion. Not only are those non-conventional uses confined to special occurrences and not very widespread

in conversation, but they even seem to be possible only when operating against a background of standing linguistic conventions.[9] First, in almost all cases, even if to varying degrees, the figuring out of the intended first meaning seems only possible thanks to the prevalence of standards in the linguistic context of the misused expression, be it simply the rest of the sentence or even more than that, some stretch of the ongoing conversation. How would the interpreter discern *monotony* in "monogamy" were it not for the regular interpretation of "We need a few laughs to break up the ..." (Davidson 1986, p. 90). Besides, there is often also some form of proximity or association either between the expression actually used and the expression standardly used to mean the intended meaning – typically a suggestive sound resemblance as with "epitaphs" and "epithets" – or even between the intended meaning and the meaning standardly associated with the expression employed – as when Davidson conceives the possibility of successfully using "Water!" to mean "Fire!" (Davidson 1986, p. 89).

The examples offered do not support the general non-necessity of conventions, but only a very local form of it. However, in this case – more than with *sufficiency* – Davidson frequently writes in a way that seems indeed to reveal his intention of holding and defending their complete dispensability. Already in 1982 he states his question in the most general terms:

> The question is delicate because it concerns not the truth of the claim that speech is convention-bound, but the importance and role of convention in speech. The issue may be put counterfactually: could there be communication by language without convention?
>
> (Davidson 1982, p. 265)

To this, of course, he answers positively, and still with no hint that he means to confine his conclusions to mere local exceptions:

> In conclusion, then, I want to urge that linguistic communication does not require, though it very often makes use of, rule-governed repetition; and in that case, convention does not help explain what is basic to linguistic communication, though it may describe a usual, though contingent, feature.
>
> (Davidson 1982, pp. 279–280)[10]

In this section I will consider a line of thought purporting to reach beyond the examples and sustain the stronger claim – that conventions are unnecessary – in its stronger and general form. I'll refer to it as "the argument from radical

---

9 Cf. Camp (2016, pp. 127–128).

10 We find the same tone in the other papers on the topic. See, for instance, Davidson (1994, p. 119).

interpretability".[11] It is a tempting argument in favor of general non-necessity, and one that is arguably discernible in Davidson's work. In any case, I'm convinced that the argument fails, and I'll show why.[12]

Davidson starts by invoking a distinction between ideal conditions and assertions of principle, on the one hand, and matters of fact and practical considerations, on the other. He uses it to characterize the role of conventions and knowledge of conventions in linguistic practices, affirming that, instead of being essential to interpretation and communication, they are merely convenient, frequently employed but ultimately extraneous to the task. He describes knowledge of such conventions as "a practical crutch to interpretation" (Davidson 1982, p. 279), indispensable in practice, "but a crutch which, under optimum conditions for communication, we can in the end throw away, and could in theory have done without from the start" (Davidson 1982, p. 279).

The first obstacle emerges here. I find the applicability of the distinction to the discussion at hand much less clear than it might appear at first. Practical considerations are of the essence when the question is precisely what is necessary for two limited creatures to reach verbal understanding. The finite nature of the participants is of central importance here, not an accessory problem. There might be some room to negotiate which constraints to reckon and which to leave aside, what is really indispensable and what is simply hard to do without, but any particular conclusion would require detailed discussion and justification, not just a vague reference to "optimum conditions". What would be *optimum*? To have a lot of time and resources? To be a very smart and lucky interpreter? Surely omniscience and telepathy are out of the question.

Ultimately, I will conclude that Davidson doesn't manage to put forward any realistic setting or context in which meaning conventions could fairly be

---

11 Various authors have identified versions of this argument, for instance, Lepore and Ludwig (2005, p. 277 ff.), Bar-On (2016, p. 57), Camp (2016, p. 121).

12 I wish to stress that I'm here committing solely to the investigation of *one* particular line of thought in defense of the strong dispensability thesis, the one that I found most prominent and promising. I'm not offering an exhaustive inspection of all available or tried arguments, nor am I claiming a final verdict on this question. Other arguments have been sketched over the years. I distinguish one in particular that was invoked by one of the reviewers; we might describe it as *the argument from indeterminacy of interpretation*. This argument made some important appearances in a number of texts from the late eighties and early nineties. It was (in some way) suggested by Ramberg (1989), closely inspected and criticized by Bar-On and Risjord (1992), and again rejected, on different grounds, by Pietroski (1994). Perhaps more could still be said about or extracted from this argument. I don't find it promising. In particular, I'm convinced by the reasons uncovered both by Bar-On and Risjord and by Pietroski that recommend its abandonment.

described as mere *practical crutches,* where speakers could plausibly reach verbal communication without relying on extensive knowledge of regularities in each others' use of language. I will argue that the prime candidate for such a setting, the fundamental situation of radical interpretation (Davidson 1973), reveals itself not as *convention free* but rather as the very enactment of a process of initiation of the interpreter into a series of conventions of language that he will end up sharing with the speaker.

Davidson originally introduced the idea of radical interpretation as the task or process of interpretation from scratch, that is, from a position of total ignorance of the speaker's language, which makes no use of dictionaries, bilingual intermediaries, or any other shortcut into the relevant meaning conventions and standards. He sketched a method for simultaneously breaking into the speaker's meanings and thoughts, starting only from her observable behavior: her moves, sounds and interactions with the environment and with other agents. In this, the interpreter is guided by a single principle, the Principle of Charity, that, at its most general, simply demands that the thoughts and meanings arrived at be such as to reveal the speaker as a rational being.

Famously, reflection on this idealized interpretative situation has a very fundamental place in Davidson's philosophy. References to it are also present in Davidson's thoughts on conventions, especially in "Communication and Convention". There he justifies the practical convenience of meaning conventions simply with our not having "the time, patience, or opportunity to evolve a new theory of interpretation for each speaker" (Davidson 1982, p. 278),[13] and implies that falling for a conventionalist model of linguistic communication requires having "[t]he fact that radical interpretation is so commonplace… hidden from us" (Davidson 1982, p. 279). However, here, Davidson is not necessarily referring to the fundamental setting described above. Plausibly, in this context he means not interpretation from scratch but simply interpretation starting from some position of less than perfect prior knowledge of the speaker's language. The crucial similarity is the general form of both investigations, the fact that, at whatever the stage, each interpretative move responds ultimately only to the Principle of Charity. We are here talking (flagrantly) about the same type of exercise in rationality and rationalization that was introduced in Section 4 above.

Can radical interpretation be used to create convention-free linguistic communication? Or, in other words: Would the radical interpretability of

---

13 This point resurfaces in Davidson (1994, p. 118).

meanings prove the general non-necessity hypothesis? The answer is negative in both senses of the enterprise.

Let us start with radical interpretation in the *mitigated* sense. Here, the negative answer follows almost immediately. In these cases we are assuming that some knowledge of the language is in place and is being relied upon to inquire into what is not already known. This form of possession and employment of shared knowledge effectively amounts to participation in meaning conventions. What we have here is the kind of examples inspected at the start of this section. There we stressed the obvious, that is, how each of these particular instances depended on a background of standing linguistic conventions.

The second scenario seems much more promising. In the fundamental setting, the radical interpreter starts from a position of total ignorance of all conventions. In this, of course, they are not unlike each of us, speakers, born unlearned. Should we not be allowed to conclude that conventions are generally unnecessary from the eventual success of the radical interpreter? The answer is no. The flaw in the inference is that it misses the fact that the process of arriving at such an understanding consists precisely in the learning or establishing of linguistic conventions.

Lepore and Ludwig come close this point when they explain that the fact that "field linguists can break into alien languages" (Lepore and Ludwig 2005, pp. 279–280) does not prove non-necessity because "they do so by figuring out the regularities in the uses of words by their subjects, which is a matter of learning which conventions govern their words in their linguistic community" (Lepore and Ludwig 2005, pp. 279–280). But there's more to it than this. The problem is not that, as a matter of fact, field linguists and children learn languages by learning conventional uses already established in the relevant communities. Radical interpretation could conceivably bring interlocutors to coordinate on a radically new language without any ties to prior conventions. The problem is that even those forged, new uses would be no less conventional, in the relevant Lewisian sense. Let us see, in more detail, how.

In the early beginnings of interpretation there is no alternative but for the speaker to use her words in a regular way.[14] "The best the speaker can do is to be

---

[14] Kathrin Glüer acknowledges that "a speaker who does not use his words with a certain regularity would not seem to be radically interpretable" (Glüer 2011, p. 110; 2013, p. 353), but still finds no "deep tension" between Davidson's views on radical interpretation and his anti-conventionalism. A reconciliation plan is suggested when she describes radical interpretation as a "limiting case" (ibid.). This argument fails, at least in the present context. An appeal to the exceptional character of this interpretative situation cannot help here because we started, pre-

interpretable, that is, to use a finite supply of distinguishable sounds applied consistently to objects and situations he believes are apparent to his hearer" (Davidson 1984, p. 13). As for the interpreter, he as well must build and test his interpretative hypotheses upon the assumption that the speaker is conducting herself in precisely that way. Let us simplify the community to include only the speaker and the interpreter – it makes no difference to my argument if interpreter and speaker are establishing new conventions of their own or simply seeking to coordinate on an already established and more widely shared practice. Take a very rudimentary example, for instance, the regularity *R of* using "There is a rabbit" to mean *there is a rabbit* – that is, the regularity of uttering "There is a rabbit" to mean *there is a rabbit*, and of taking someone else's utterance of "There is a rabbit" to mean *there is a rabbit*. First, only the speaker is conforming to *R*, while the interpreter is assuming that there is some regularity but is still trying to figure out what. He muses that it will have something to do with rabbits, or maybe animals, or maybe game or meat. Only new observations can confirm with reasonable certainty that the right meaning is *there's a rabbit*, and, by then, conformity to *R* will already be common to interpreter and speaker. Even if the interpreter, for some reason, were not himself uttering the learned expression, he would still be conforming to *R* just by taking the speaker's utterances to mean the relevant meaning.

Furthermore, all this already gets us more than simple shared conformity. Regularity and repetition are vital in this process. But once we grant that regularity on the part of the speaker and the correspondent knowledge on the part of the interpreter must be in place, it is relatively easy to see that the remaining conditions for there to be a convention, in the Lewisian sense, are also satisfied.

Let us first consider *common knowledge*. The transition into competence with some linguistic expression will likely be a very diffuse event. However, from the moment where there is solid enough understanding for some expression, we can be confident that interpreter and speaker will be sharing not only conformity to the relevant regularity but also common knowledge that they do. To begin with, each must know of his or her own case that he or she is conforming to R. Next, by now we can also be sure that the interpreter knows of the regularity on the

---

cisely, by establishing that non-radical interpretation cases fared no better in rendering plausible general non-necessity. It must be said, however, that this is not meant as a direct challenge to Glüer's position, because it is not clear whether she takes something as strong as *general* non-necessity to be at stake in this discussion.

speaker's part. It is only because he managed to figure that out that he himself is conforming to *R*.

There follows a slightly more challenging question: why must it be the case that the speaker reciprocates the attention? One way to defend this is to invoke Davidson's criterion for communicative success – that the speaker must manage to get her words to be understood as she *intends* them to be understood. When pressed with the *Humpty Dumpty objection* (Davidson, 1986, pp. 97–98), Davidson explains that he's not advocating the whimsical and unsuccessful style of linguistic innovation attempted by Carroll's character. A genuine intention, Davidson insists, requires a reasonable expectation of realization; therefore, a "speaker cannot … intend to mean something by what he says unless he believes his audience will interpret his words as he intends" (Davidson 1986, p. 97). This translates to our present context; our *radical* speaker is no less required to possess a reasonable belief that her words will be understood as she intends. That is, the speaker's knowledge that the interpreter is himself conforming to *R* is also a necessary condition for verbal communication in Davidson's model. Without it there simply is no interpretation, and hence no radical interpretation.

Once this minimal basis of common knowledge has been reached – and taking into account that interlocutors are rational agents genuinely interested in communication and mutual understanding, and that all the *investigation* takes place *in the open* – the following layers of the conceptual spiral become likewise accessible. Each interlocutor will then be in a condition to infer that each interlocutor knows that both are conforming to R, and that each interlocutor knows that each interlocutor knows that both are conforming to R, and so on, indefinitely.

In the discussion of common knowledge, we already touched on a further article of Lewis' analysis, namely the condition that it is the belief that others conform to R that provides everyone a good and decisive reason to conform to it. We have already noted that this is how it is with the interpreter: to figure out the regularities in the speaker's linguistic conduct so as to be able to ascribe her meanings in accordance is the proper goal of his interpretive effort. The same must also be true of the speaker. We can infer it as the best explanation for her regular conduct. It is only because she expects that her regularity will allow the interpreter, sooner or later, to interpret adequately her words that she holds on to her consistent practice. That is, it is only because she hopes and aims for the interpreter's conformity to *R* that she herself conforms to it.

The two final conditions are also fulfilled. First, once interpreter and speaker have started, there comes into place a general preference for general conformity to *R*. Second, it is also the case that there are alternatives to R that would be just

as viable, as attested by the great variety of ways, in different languages, of meaning *there is a rabbit.*

All Lewisian conditions are fulfilled. I conclude, as anticipated, that radical interpretability implies no general dispensability of conventions for linguistic communication. Instead of the promised convention-free alternative, radical interpretation would constitute the very bringing about of conventional solutions to their communicative problems.

Lepore and Ludwig try a slightly different approach. They avail themselves of a distinction evocative of the one Davidson appeals to between matters of principle and matters of fact. Ultimately they reject radical interpretability but they maintain that it would actually imply *in principle* general[15] dispensability of conventions, even if we, common agents, do not, and could not, proceed this way due to our contingent limitations. Here is how they do it. They start by affirming that to ask "whether it is in principle possible to interpret another without appeal to prior knowledge of conventions" (Lepore and Ludwig 2005, p. 278) is tantamount to asking "is there knowledge an interpreter could in principle have, leaving aside natural limitations of knowledge and perspicacity, which would enable him to correctly interpret a speaker of whom he had no prior knowledge at some given time?" (Lepore and Ludwig 2005, p. 278). This latter they find "equivalent to asking whether there are facts independent of linguistic conventions that determine what a speaker means by his words" (Lepore and Ludwig 2005, p. 279). Thus they conclude that "if Davidson's basic methodological stance on matters of meaning is correct" (Lepore and Ludwig 2005, p. 279) – that is, if speakers are radically interpretable – "then the answer is clearly affirmative" (Lepore and Ludwig 2005, p. 279).

I follow the first part of the argument. To affirm the radical interpretability of the speaker is to affirm the existence of meaning-determining facts that are accessible to interpreters with no prior knowledge of the language or of its conventions. My doubts concern what comes next. Would these facts be really "independent of linguistic conventions" (Lepore and Ludwig 2005, p. 279)? There is a sense in which they would indeed be independent: they *are accessible* to people outside those conventions and ignorant of the relevant knowledge. However, if I'm right, there is second a sense, and a crucial one, in which they would not: these facts would *not be the case* were it not for standing or emergent linguistic conventions.

---

15 They do not pause at the general versus local distinction for non-necessity – see Lepore and Ludwig (2005, p. 269) – but it is clear, in context, that they are thinking about the general.

Lepore and Ludwig seem to acknowledge only the speaker's role in meaning determination, and they simply start with the assumption that the relevant meaning-determining facts are there, publicly available in the world, waiting to be discovered. The whole question, then, becomes whether or not the interpreter is able to collect them, to which the *in principle* qualification allows a positive answer. In other words, they subtract the limited human interpreter from Davidson's picture and assume that the all the relevant facts would remain in place.

However, meaning-determining facts don't happen for no reason.[16] In particular, speakers exhibit whatever dispositions to verbal behavior they exhibit because they want to make themselves understood to some audience and expect to accomplish that by acting in some particular way. In this section, we found nothing supporting the possibility of holding a reasonable expectation of being verbally understood that did not appeal to meaning conventions: regular uses are conventional, new uses rely on a background of regular uses, radically interpretable uses aim at being conventionalized. Speakers' behaviors will reflect this state of affairs. Without standing conventions, be it in the forefront or in the background, or the prospect of new ones, speakers will simply not find the right opportunity for any linguistic move. Speakers' linguistic behavior, as much as interpreters' inquiry, depends on linguistic conventions.

6 Conclusion

In this paper I have reconsidered the role of convention in linguistic communication within Davidson's framework. Occasional examples of convention-free successful verbal exchanges are frequent enough, but the intuitive reaction is to take them as mere exceptions of limited consequence for the general understanding of linguistic communication. I followed two arguments purporting to show that the split is in fact more robust. I concluded, in the first case, in favor of the general insufficiency of meaning conventions to secure that interlocutors coordinate on the literal meaning of their utterances. I concluded, in the second case, against the general non-necessity of conventions. Linguistic practice engages speakers in a continuous exercise of vigilance and deeper interpretation; at no moment are they permitted to rest exclusively on their stock knowledge of the prevailing linguistic conventions. However, this is not enough to deprive conventions of an indispensable role in the affair. I tried to

---

16 Cf. Verheggen (2016, p. 69).

make clear that instead of competing models, conventions of meaning emerge and strive in essential interdependence with other efforts of rationality and rationalization.

Acknowledgements

References

Bar-On, Dorit (2016): "Sociality, Expression, and This Thing called Language". In: *Inquiry* 59(1), pp. 56–79.

Bar-On, Dorit/Risjord, Mark (1992): "Is There Such a Thing as a Language?" *Canadian Journal of Philosophy* 22, no. 2, pp. 163–190.

Camp, Elisabeth (2016): "Convention's Revenge: Davidson, Derangement and Dormitivity". In: *Inquiry* 59(1), pp. 113–138.

Davidson, Donald (1965): "Theories of Meaning and Learnable Languages". In: Donald Davidson (Ed.): *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, pp. 3–15.

Davidson, Donald (1967): "Truth and Meaning". In: Donald Davidson (Ed.): *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, pp. 17–36.

Davidson, Donald (1973): "Radical Interpretation". In: Donald Davidson (Ed.): *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, pp. 125–139.

Davidson, Donald (1982): "Communication and Convention". In: Donald Davidson (Ed.): *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, pp. 199–214.

Davidson, Donald (1986): "A Nice Derangement of Epitaphs". In: Donald Davidson (Ed.): *Truth, Language, and History*. Oxford: Clarendon Press, pp. 89–107.

Davidson, Donald (1992): "The Third Man". In: Donald Davidson (Ed.): *Truth, Language, and History*. Oxford: Clarendon Press, pp. 159–165.

Davidson, Donald (1993): "Locating Literary Language". In: Donald Davidson (Ed.): *Truth, Language, and History*. Oxford: Clarendon Press, pp. 167–181.

Davidson, Donald (1994): "The Social Aspect of Language". In: Donald Davidson (Ed.): *Truth, Language, and History*. Oxford: Clarendon Press, pp. 109–125.

Dummett, Michael (1986): "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking". In: Ernest Lepore (Ed.): 1986, *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell, pp. 459–476.

Green, Kathryn (2001): "Davidson's Derangement: Of the Conceptual Priority of Language". In: *Dialectica* 55(3), pp. 239–258.

Glüer, Kathryn (2011): *Donald Davidson: A Short Introduction*. New York: Oxford University Press.

Glüer, Kathryn (2013): "Convention and Meaning". In: Ernest Lepore/Kirk Ludwig (Eds.): *A Companion to Donald Davidson*, Malden, MA/Oxford: Wiley-Blackwell, pp. 339–360.

Lepore, Ernest/Ludwig, Kirk (2005): *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford University Press.

Lepore, Ernest/Stone, Matthew (2015): *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford: Oxford University Press.

Lewis, David (1969): *Convention: A Philosophical Study*. (2002) ed. Oxford: Blackwell.

Lewis, David (1975): "Language and Languages". In: David Lewis (Ed.): *Philosophical Papers: Volume I*, 1 edition, New York: Oxford University Press, pp. 163–188.

Nolan, Daniel (2005): *David Lewis*. Chesham: Routledge.

Pietroski, Paul (1994): "A Defense of Derangement". In: *Canadian Journal of Philosophy* 24(1), pp. 95–118.

Ramberg, Bjørn (1989): Donald Davidson's Philosophy of Language: An Introduction. Oxford: Blackwell.

Reimer, Marga (2004): "What Malaprops Mean: A Reply to Donald Davidson". In: *Erkenntnis* 60, pp. 317–334.

Verheggen, Claudine (2016): "From Triangulation to Semantic Externalism". In: Robert H. Myers/Claudine Verheggen, *Donald Davidson's Triangulation Argument: A Philosophical Inquiry*. London: Routledge, pp. 65–94.

Tadeusz Ciecierski
Reflective Meta-attitudes and (In)Compatibilism

Abstract: This paper contains a defense of the orthodox view that knowledge entails belief. I begin with distinguishing two ways in which one can deny the knowledge-belief entailment claim. The first is incompatibilism, which claims that co-occurrence of knowledge and belief is impossible. The second is compatibilism, which claims that that co-occurrence of knowledge and belief is a contingent matter. Next, I present intuitive arguments against both kinds of denials. The arguments apply to agents that are capable of entertaining reflective meta-belief and reflective meta-knowledge for every belief or knowledge state they have. I close the paper with a discussion of some possible objections to the presented arguments. The first is the problem of the interpretation of apparent "knowledge without belief" scenarios that enables the keeping of the contrast between them, and then also both "knowledge with belief" and "belief without knowledge" scenarios. Firstly, I consider the class of cases of apparent "knowledge without belief" scenarios encompassing situations in which one is disposed to assent that 'I do not only believe that: I know it'. Secondly, I consider the class of cases of apparent "knowledge without belief" scenarios embracing, for instance, the classical Radford example in the version presented recently by Myers-Schulz and Schwitzgebel. The second problem concerns the commitment to epistemic logics. The third is the problem that the arguments appeals to reflective agents who have the disposition to form meta-attitudes, and one may still claim that compatibilism is true of agents that are not in that group (young children, some or all animals, etc.). The fourth problem is the commitment to attempts to provide sufficient and necessary conditions for the application of the concept of knowledge.[1]

Keywords: knowledge, belief, higher order attitudes, compatibilism, incompatibilism

1 Introduction

As Jakko Hintikka described the matter in his seminal work: "It is indeed reasonable to assume that whatever one *actively* knows one also *actively* believes" (Hintikka 1962, p. 50). Call this a traditional picture of the knowledge-belief relationship. According to this traditional picture, knowledge entails belief. After Gettier (1963), it became part of common philosophical knowledge that the idea of knowledge that comprises this traditional view must be wrong. Most philosophers find the source of idea's weakness in the notion of justification. Replacing this notion with a concept that appeals to contextually relevant standards (DeRose 1992) or counterfactual dependency of belief on circumstances of evaluation (Nozick 1981) are among the most popular strategies for dealing with its weaknesses. Other strategies attempt to supplement the idea of knowledge with the fourth condition that is not met in Russell-Gettier cases (cf. the "no false grounds" view of Feit and Cullison, 2011). There remains, however, a smaller group of philosophers who look for a remedy elsewhere – they assent to the very possibility of knowledge without belief[2] and therefore question the traditional picture of the knowledge-belief relationship. Below I will present considerations that support the claim that they must be wrong, at least if among cognitive agents there are subjects capable of entertaining reflective meta-beliefs and reflective meta-knowledge states. Such an argument may be important since, as some critics have complained, the dominant strategy of arguments for the correctness of the traditional picture is a 'wait-for-counter-examples' approach (Myers-Schulz and Schwitzgebel 2013).

The structure of this chapter is as follows. In the first section I shall contrast two ways in which one can deny the knowledge-belief entailment claim. In the second and the third sections, I shall present intuitive arguments against both kinds of denial. In the fourth and final section I shall address some possible objections against the arguments from sections two and three.

2 Compatibilism and Incompatibilism

Let us borrow a standard notation from epistemic and doxastic logic (I shall say more about the commitment to such logics in the final section of the paper). Let us, therefore, abbreviate 'x knows that p' as '$K_x p$' and 'x believes that p' as '$B_x p$'.

---

2 Yet an even smaller group of philosophers denies the factivity of knowledge. I do not discuss this issue in this paper.

In this notation, the entailment thesis may be formulated in the following manner:

$$(KB) \quad \Box \forall x \, (K_x p \Rightarrow B_x p)$$

which (assuming the converse Barcan formula) immediately gives the following claim about actuality:

$$(AKB) \quad \forall x \, \Box \, (K_x p \Rightarrow B_x p)$$

The denial of (KB) is:

$$(NKB) \, \Diamond \, \exists x \, (K_x p \wedge \neg B_x p)$$

which, obviously, does not have to be exemplified by any actual cognitive agent (merely a possible one suffices). As such, (NKB) does not say much about the source of the possibility of knowledge without belief but, as it is equivalent to (if the accessibility relation is serial):

$$(NKB\star) \, [\Diamond \, \exists x \, (K_x p \wedge \neg B_x p) \wedge \Diamond \, \exists x \, (K_x p \wedge B_x p)] \vee \neg \, \Diamond \, \exists x \, (K_x p \wedge B_x p)$$

it embraces two distinct claims about the relationship between knowledge and belief. The first one, which I will refer to as *compatibilism*, states that:

$$(COM) \, \Diamond \, \exists x \, (K_x p \wedge \neg B_x p) \wedge \Diamond \, \exists x \, (K_x p \wedge B_x p)$$

while the other, which I will refer to as *incompatibilism*, is:

$$(INCOM) \, \neg \, \Diamond \, \exists x \, (K_x p \wedge B_x p)$$

Compatibilism is the claim that the co-occurrence of knowledge and belief is a purely contingent matter, while incompatibilism states that co-occurrence is impossible.

Roughly speaking, incompatibilism is a radical thesis, and it is difficult to see how one can find reasons for believing it. It goes without saying that we may utter a sentence *p* and, for instance, truly say afterwards: "I believe and know that to be true". If one takes incompatibilism to be true, one has to deny the intuitive truthfulness of such assertions. This clearly requires an argument, as well as a theory, that explains why we should treat them as perfectly correct and possibly true. Furthermore, one may note that if incompatibilism is true, then – since in

such a case, belief that p would entail ignorance about p and knowledge would entail the lack of belief – we have to face the fact that particular instances of conjunctions '$B_x p \wedge \neg K_x p$' and '$K_x p \wedge \neg B_x p$' are redundant, as is generally true of conjunctions of the form '$p \wedge q$' in cases where p entails q or vice versa.[3] We must likewise recognize that it is difficult to recognize the redundancy or triviality of such utterances.

It may be worth noting, however, that there exists an interesting connection between incompatibilism and skepticism about the external world. Bertrand Russell noted once that between the content of the mental state and its objective (i.e. the fact that makes it true or false):

> (...) there is sometimes a very wide gulf, for example in the case of 'Caesar crossed the Rubicon.' This gulf may, when it is first perceived, give us a feeling that we cannot really 'know' anything about the outer world. All we can 'know,' it may be said, is what is now in our thoughts. If Caesar and the Rubicon cannot be bodily in our thoughts, it might seem as though we must remain cut off from knowledge of them.
>
> (Russell 1992, p. 234)

As I understand him, Russell interprets external world skepticism (he is not taking this position, of course) as defending a version of restricted incompatibilism: beliefs about the external world are not (and cannot be) knowledge states, and (equivalently) knowledge states about the external world cannot be belief states. This is usually interpreted as the claim that there are no knowledge states (about the external world) at all, but Russell's description of the theory fits a different interpretation very well: knowledge states are not a species of belief states.[4]

I must admit that I know of no author who firmly defends incompatibilism (external worlds skeptics aside). The only person who defended a view *relatively* close to it is Raziel Abelson (Abelson 1968), who claimed that *in cases of first-person sentences* it is pragmatically inappropriate to utter a belief sentence and the corresponding knowledge sentence, as the correctness of the former pragmatically entails incorrectness of the latter. In his analysis, Abelson was not making use of the notion of implicature, but one may interpret his suggestion as the claim that, in normal contexts, a belief that the speaker does not know that p is a scalar

---

3 I would like to thank Tomasz Puczyłowski for stressing this point in discussion.

4 Does this interpretation extend to radical skepticism? Since Russell explicitly relies on the distinction between knowledge of our mental states and knowledge of external facts, the reasons he describes seem inapplicable to skepticism, appealing as they do to Cartesian-Demon-like scenarios (excluding states that are constitutive of *cogito* if one takes them to be directly knowable). However, the radical skeptic clearly defends incompatibilism too, as the radical skeptic thinks that the Russellian gulf extends to every belief.

implicature of the utterance 'I believe that *p*'. Since this implicature can be cancelled in particular contexts, utterances like 'I believe and know that to be true' are perfectly correct and possibly true in such contexts. Interpreted in such a way, Abelson's theory can be reconciled with both compatibilism and the traditional picture of the knowledge-belief relationship.

Compatibilism, on the other hand, has been defended by several prominent thinkers. One can mention here, for instance, Radford (Radford 1966), Black (Black 1971), Mannison (Mannison 1976), and Lewis (Lewis 1996). Some recent results in experimental philosophy (cf. Myers-Schulz and Schwitzgebel 2013) may also be interpreted as supporting compatibilism. One of the arguments for compatibilism postulates a strict symmetry between cases of intuitively correct utterances of sentences like 'A believes and knows that to be true' and intuitively correct utterances of sentences like 'A knows this, but s/he doesn't believe it'. Consider, for instance, the following example from Myers-Schulz and Schwitzgebel (which is a variation of the example discussed originally by Radford[5]):

> Kate has spent many hours studying for her history exam. She is now in class taking the exam. Everything is going quite well, until she comes to the final question. It reads, "In what year did Queen Elizabeth die?" Kate has reviewed this date many times. She even recited the date to a friend just a few hours earlier. So, when Kate sees that this is the last question she feels relieved. She confidently looks down at the blank space, waiting to recollect the answer. But before she can remember it, the teacher interrupts and announces, "All right, the class session is almost over. You have one more minute to finalize your answers." Kate's demeanor suddenly changes. She glances up at the clock, now flustered and worried. "Oh, no. I can't perform well under this kind of pressure." Her grip tightens around her pencil. She strains to recall the answer, but nothing comes to her. She quickly loses confidence. "I suppose I'll just have to guess the answer," she says to herself. With a sigh of disappointment, she decides to write "1603" in the blank space. This is, in fact, the correct answer.

The scenario has two versions: one with the summarizing question "Did Kate know that Queen Elizabeth died in 1603?" and the other with the question "Did Kate believe that Queen Elizabeth died in 1603?" The answers showed a statistically significant difference in positive replies to both questions. This might be taken as evidence of a case of knowledge without belief. If this is correct, the argument obtains, since we take cases of 'A believes and knows that to be true' as instances of correct and possibly true utterances, exactly the same has to be said

---

5 A similar scenario is mentioned by Lewis: "I even allow knowledge without belief, as in the case of the timid student who knows the answer but has no confidence that he has it right, and so does not believe what he knows" (Lewis 1996, p. 556).

of utterances like 'Kate knows that Queen Elizabeth died in 1603, but she doesn't believe that Queen Elizabeth died in 1603'. Hence, proponents of the standard view have to explain why cases of the latter kind are not in fact similar to cases of the former kind. I shall come back to this challenge in section three below.

3 An intuitive argument against incompatibilism

In spite of the fact that incompatibilism is a radically unattractive view that has unwelcome consequences, it might be useful to have a direct argument against it. Here is one that seems convincing. Assume (INCOM) and the factivity of knowledge. This gives us (for a particular agent a):

$$\Box\, [\mathrm{K_a}p \Rightarrow (p \wedge \neg \mathrm{B_a}p)]$$

which, if p is any sentence stating that '$\mathrm{B_a}q$,' provides:

$$\Box\, [\mathrm{K_a}\mathrm{B_a}q \Rightarrow (\mathrm{B_a}q \wedge \neg \mathrm{B_a}\mathrm{B_a}q)]$$

Now suppose, uncontroversially, that:

$$\Diamond\, [\mathrm{K_a}\mathrm{B_a}q \wedge (\mathrm{B_a}q \Leftrightarrow \mathrm{B_a}\mathrm{B_a}q)]$$

i.e. that there is a possible scenario in which person a knows that s/he believes something and that (in that scenario) s/he believes that s/he believes what s/he believes. Now it follows immediately that:

$$\Diamond(\mathrm{B_a}\mathrm{B_a}q \wedge \neg \mathrm{B_a}\mathrm{B_a}q).$$

That is, in that in this scenario s/he is simultaneously holding and not holding a particular meta-belief. Allowing such an impossibility constitutes, in my opinion, a decisive argument against incompatibilism.

4 An intuitive argument for the traditional picture

In order to argue convincingly for (KB), the following condition has to be met: in the argument, neither the (KB) nor the corresponding inference rule ("$\mathrm{K_x}A$ entails $\mathrm{B_x}A$") may be used. However, this does not apply to every *instance* of (KB) if we

have *independent* reasons to think that a particular instance is justified. Consider, for instance, the following statement:

$$\text{(MKB)} \ \Box \forall x \ (K_x K_x p \Rightarrow K_x B_x p)$$

Suppose that you have reflective meta-knowledge about your knowledge that *p*. This requires that a certain repertoire of epistemic constraint regarding your attitude towards *p* – *p* has to be true as well, in addition to other conditions that knowledge has to meet. The same applies to reflective meta-knowledge that requires a certain repertoire of epistemic constraint regarding one's attitude towards *knowledge that p* – it has to be the case that one knows that *p* as well, and that other constraints on knowledge hold true. As noted above, the nature of such additional conditions remains controversial. However, one may argue that no matter what the conditions are, they are roughly met in the case of knowledge about our own mental states and, in particular, in the case of knowledge about our knowledge. If you think you know that you know something, then, if it is, in fact, true that you know this, then your reflective grounds for thinking this ensure that the other constraints imposed on knowledge are met. And, if they are met (e.g. if you are justified in thinking that you know that you know that *p*, or if this thought tracks the truth, or if no false grounds for thinking that are involved), then you clearly cannot know that you disbelieve that *p*. So, one may have independent reasons for believing (MKB) even if one does not accept (KB).

The simple reasoning I present below applies to agents that are capable of entertaining reflective meta-belief and reflective meta-knowledge[6] for every belief or knowledge state they have. Thus, a version of the traditional picture we are going to argue for looks like this:

$$\text{(KB⋆)} \ \Box \forall x \ [x \text{ is capable of having reflective meta-attitudes} \Rightarrow (K_x p \Rightarrow B_x p)]$$

In fact, it is sufficient even to assume that they are capable of entertaining reflective meta-belief and reflective meta-knowledge for every belief or knowledge state *that is not a meta-belief or meta-knowledge state*. This may be important, as one may doubt if allowing meta-beliefs about meta-beliefs, meta-knowledge about meta-knowledge, meta-beliefs about meta-knowledge, or meta-knowledge about meta-beliefs makes sense.[7] Below I shall argue for (KB⋆) directly, but the

---

6 Occasionally I shall speak of meta-attitudes here.

7 I am not claiming that it does not. The point is simply that we do not have to assume this in does in order for the argument to work. More about this below – see: *Problem 2* in section four.

only assumption the argument requires is that the agents have a disposition to entertain attitudes that are not themselves attitudes about attitudes.

So here is the first simple, albeit controversial, argument. Assume that one knows that p:

$$(1)\, \mathrm{K}_a p$$

Now, one may assume that for reflective agents the following holds as well:[8]

$$(2)\, \Box \forall x\, (\mathrm{K}_x p \Rightarrow \mathrm{K}_x \mathrm{K}_x p)$$

This is, of course, the (in)famous KK principle, and we have more to say about its relevance and justification (see below). (2) and (1) immediately provide:

$$(3)\, \mathrm{K}_a \mathrm{K}_a p$$

and this (assuming (MKB)) gives:

$$(4)\, \mathrm{K}_a \mathrm{B}_a p$$

which immediately entails (assuming the factivity of knowledge):

$$(5)\, \mathrm{B}_a p$$

Since all the additional assumptions used in the argument (that is: (2), (MKB), and the factivity of knowledge) are necessarily true, it follows that (KB*) is true. The argument is simple, intuitive, and presupposes neither (KB) nor the corresponding inference rule. As such, it appears a good candidate for a reply to those philosophers who complain about the 'wait-for-counterexamples' strategy of arguing for (KB). Its strength depends, of course, on the legitimacy of the assumptions used.

The last of the assumptions is factivity (in our version: '$\Box \forall x\, (\mathrm{K}_x p \Rightarrow p)$'). I take this assumption to be uncontroversial (the best argument in its favor might be the nonexistence of convincing counterexamples). Another assumption is (2), or the (in)famous KK principle. This assumption is controversial. It has been defended, for instance, by Hintikka (1962, 1970) and criticized by Williamson (2000)

---

8 This is the only principle regarding the reflective meta-attitudes that one has to assume in this version of the argument.

(see Hemp (ONLINE) or a useful overview of the controversy – Hemp suggests that *both* Hintikka and Williamson may be correct, as they are describing distinct notions of knowledge). I do not aim to resolve the controversy here. On the contrary, I believe that the controversy over the KK principle clearly indicates the need for an alternative argument.[9]

I believe that the following principle of "introspection" may help us here:

$$(6)\Box\forall x\,(B_x p \Leftrightarrow B_x B_x p)$$

The principle seems to apply to reflective agents that are capable of having reflective meta-beliefs. (It is, for instance, a modalized version of one of the axioms of the logic LB (the logic of conscious beliefs), cf. Tokarz, 1990.) Note that the fact (noticed by numerous philosophers) that one may be mistaken if one is in the belief-that-p state is not a counterexample to (6): even if I am wrong that I am brave enough to do a parachute jump (I have never been in the situation in which my disposition or indisposition to jump has been tested), this proves at most that *I do not know* if I really believe something. Another statement that should be true of reflective agents is as follows:

$$(7)\ \Box\forall x\,(K_x p \Rightarrow B_x[(K_x p \wedge B_x p) \vee (\neg K_x p \wedge B_x p)])$$

(7) states that knowledge implies the belief that either one knows while believing or one believes without knowing. This is a simple formulation of the view that knowledge states and belief states are subjectively indistinguishable. I think that (7) is a specific case of a more general principle stating that if two attitudes of the agent differ at most in external (i.e. subject-independent) credibility, then the subject has to believe that s/he is at least in the state of having fewer strict and demanding credibility standards.

Now (7) (by propositional logic and distribution of belief over conjunction) provides:

$$(8)\ \Box\forall x\,(K_x p \Rightarrow B_x B_x p)$$

---

9 The minimal conclusion that can be derived is that accepting the KK principle should provide one good reason for believing in the traditional picture.

and this (by (6)) directly entails (KB*). Just as in the previous case, the argument assumes neither (KB) nor the corresponding inference rule. The only required additional assumptions are (6) and (7) which, I think, can be independently justified in the case of reflective and conscious agents.

5 Problems

The argument cannot be dismissed, however. I shall concentrate on four possible objections that one could raise against the considerations presented.

Problem 1. The first problem we have to face is the interpretation of apparent "knowledge without belief" scenarios, while keeping the contrast between such scenarios and "knowledge with belief" and "belief without knowledge" cases.

Reply. The first class of cases of apparent "knowledge without belief" scenarios encompasses situations in which one is disposed to assent to '*I do not only believe that: I know it*' (or assent to a sentence that, in a given context, can be treated as equivalent to such a locution). However, this hardly produces a counterexample to (KB), as uttering 'I do not only believe that: I know it' seems to be just a way of saying that one is not in a state of believing without knowing but rather in a state of *believing and knowing*. Moreover, this fits the implicature analysis sketched above: by assenting to 'I do not only believe that: I know it' one simply cancels the scalar implicature of regular belief sentences[10] (one may note immediately that such cases are instances of "knowledge with belief" situations). It goes without saying that it is hard to see how the implicature strategy might be applicable to the "belief without knowledge" scenario.

　　The second class of cases of apparent "knowledge without belief" embraces, for instance, the Radford example in the version presented by Myers-Schulz and Schwitzgebel. The scenario is interesting partially because (in contrast with the examples of the first kind) the implicature strategy seems to be inapplicable to it. As such, I think, it indicates that it is incorrect to claim that:

---

10 This corresponds well to what certain other authors think of such situations (cf. Lehrer 1990).

It makes sense to say, 'I do not believe that; I know it', not because it is logically inconsistent to say that a woman believes what she knows but rather because this is an emphatic way of saying, 'I do not only believe that: I know it'

(Lehrer 1990, p 27).

The reason is that not every utterance of 'I do not believe that; I know it' can be taken as tantamount to 'I do not only believe that: I know it'. In the Radford-Myers-Schulz-Schwitzgebel scenario we clearly do not want to say that Kate does not only believe that Queen Elizabeth died in 1603. That being said, the interpretation of this (and other) empirical considerations used against the entailment thesis is very controversial.

The conclusions of Myers-Schulz and Schwitzgebel have been questioned by other authors who have claimed that experimental considerations in fact support the view that knowledge entails *dispositional belief* (Rose and Schaffer 2013) or that knowledge entails the so-called *thin belief* [11] (Buckwalter, Rose and Turri 2015). Both views are taken by the authors to be the intended understanding of the entailment thesis. I generally agree with the criticism presented in the two aforementioned papers. This agreement notwithstanding, I would like to indicate briefly another possible weakness of the Myers-Schulz and Schwitzgebel interpretation of the Radford scenario (or similar scenarios).

In actual fact, I want to claim that there is a sense in which *we* may *say* that Kate knows and disbelieves that Queen Elizabeth died in 1603. This, however, comes at a cost: if we grant this, *we* must claim that there is an unexplained asymmetry between the third- and first-person cases of attitude attributions. Consider the following version of the original story:

You have spent many hours studying for your history exam and you are now in class taking the exam. Everything is going quite well, until you come to the final question. It reads, "In what year did Queen Elizabeth die?" You have reviewed this date many times... (then the story continuous as in the original scenario). After the exam you meet a friend who asks, "What was your answer to the question: In what year did Queen Elizabeth die?" You reply: "1603." The friend applauds: "That is correct! I am glad that you knew it!"

What will be your *honest* reply to this comment? I have no doubt that it will be along the lines of: "No, I didn't. I had to guess and I was lucky. I knew it before the exam but when the teacher announced that the class session was almost over

---

11 Roughly speaking, *thin belief* that *p* is the belief that requires nothing more than regarding *p* as true. The other category of *thick belief* requires additionally that emotional (and motivational) element is present. Buckwalter and Rose think that folk psychology distinguishes clearly two notions of belief that correspond to this distinction.

I simply couldn't recall the date."[12] (The answer to the belief question is going to be straightforwardly negative.) Now, the only difference between the scenarios is that we have switched the third person to the *first person*. It follows that, in fact, here we do not have a *knowledge without belief* scenario but rather an *ignorance with disbelief scenario*. I must admit that I have no idea how proponents of compatibilism are going to explain the asymmetry between the two cases. I prefer to think of both scenarios as analogous and the replies that confirm the hypothesis endorsed by compatibilists as incorrect. Compatibilism seems to be guilty of a one-sided diet that "nourishes one's thinking with only one kind of example" (*Philosophical Investigations*, §593).[13]

Problem 2: The considerations presented above seem to be committed to (one or other) version of epistemic and doxastic logic. But the very idea of such logic is highly problematic, as "(…) there are knowers and knowledge constituting counterexamples to every extant theorem (…) of every epistemic logic" (Hocutt 1972, p. 435). (The same might be said, of course, of believers and doxastic logic.)

Reply: Actually, the argument remains correct no matter what stance towards epistemic or doxastic logic one takes (the author, for instance, is very sympathetic towards Hocutt's skepticism). The only assumption one has to endorse is that of the minimal rationality of cognitive agents. By "minimal rationality" I mean the following constraint: if a believes that p and if q follows from p, then if a had been informed about the structure and content of the correct argument from p to q, and if s/he grasps the argument (which requires, among other things, that s/he is capable of grasping q), s/he would have added q to her/his beliefs (e.g. if s/he had believed not-q before, she would have given up this belief).[14]

---

12 Experimental philosophers might want to test this (I personally do not think it is worth testing). Probably, the scenario in which the summarizing question is: "Would the reply 'Yes, I knew it' be the honest answer to your friend's comment?" is going to be the most empirically valuable.
13 Let me add that neither Rose and Schaffer, nor Buckwalter and Rose note the first- and third-person asymmetry problem, they are also not interested in agents capable of having higher order attitudes. Both issues, I think, may happen to be of some relevance to the study of differences in perspective and "projective use of language" in attitude attribution studies (the relevance of perspective and projectiveness is stressed by Rose and Schaffer).
14 The idea is, of course, taken from Hintikka who writes (explaining the notions of defensibility and indefensibility):

"(…) suppose that a man says to you, 'I know that p but I don't know whether q', and suppose that p can be shown to entail q logically by means of some argument he would be willing to accept. Then you can point out to him that what he says he does not know is

For the sake of illustration, consider the principle (6) ('$\Box \forall x\, (B_x p \Leftrightarrow B_x B_x p)$'). One of the obvious cases against it (and similar rules) appeals to iteration and the finiteness of cognitive agents (cf. Rescher 2005, p. 22): each cognitive agent seems to have a limited attentiveness that excludes the possibility of being disposed to sufficiently complex meta-attitudes (exceeding the ability to have tenth-order thoughts, let us say). This, however, might be represented as just an additional individual constraint imposed on an agent's capacity to grasp certain propositions.[15] The constraint is applicable both to cases in which you consider a certain embedding of attitudes as a possible object of direct (i.e. non-inferential) awareness, and to cases where you consider it a possible consequence of your first-order belief that p. One may additionally note that our considerations appealed to the embedding of attitudes of minimal complexity only.

Problem 3: The argument appeals to reflective agents who have the disposition to form meta-attitudes. One may still claim that compatibilism is true of agents that are not in that group (young children, some or all animals, etc.)

Reply: Yes, the argument does not apply to such unreflective agents. However (although this remains a matter of empirical controversy), there are reasons to believe (*vide* the false-belief task and its interpretations) that agents who do not have such abilities project their own internal states onto others. The question now is what argument proponents of incompatibilism are going to provide here in support of their view. They probably cannot follow the empirical route of Myers-Schulz and Schwitzgebel because, as we noted above, there is an asymmetry between the first- and third-person scenarios here. Of course, I do not want to claim that this proves the correctness of (KB) for unreflective agents (that would

---

already implicit in what he claims he knows. If your argument is valid, it is irrational for our man to persist in saying that he does not know whether q is the case. If he is reasonable, you can thus persuade him to retract one of his statements without imparting to him any fresh information beyond certain logical relationships (the rules of which he is assumed to have mastered from the beginning). You have done this by pointing out to him that he would have come to know that q all by himself if he had followed the consequences of what he already knew far enough"

(Hintikka 1962, p. 31).

15 In epistemic logic the constraint might be imposed purely syntactically: 'If p is a formula, then $B_x p$ is a formula only if p contains no more than n iterations of B.' n might be different for distinct cognitive agents. If n is zero, then one may say that the 'logic' applies to a non-reflective agent.

be a case of *argumentum ad ignorantiam*). Nonetheless, since compatibilism is far from being an intuitive and obvious claim (both for reflective and unreflective agents), it is clearly in need of additional rationale.

Problem 4: By defending the (KB) thesis the proponent of the argument is committed to attempting to give sufficient and necessary conditions for the application of the concept of knowledge. Such attempts to define knowledge are known to be theoretical dead-ends.

Reply: I am not committed to such definitional endeavors. What is defended is a claim to the effect that *had such definitions been possible, they would have had to contain the (KB) claim* (or a clause that entails it). If they are not possible, I am very happy to remain with the claim that believing is a necessary condition of knowing.

6 Conclusion

I conclude that there are reasons to think that the traditional picture is correct. Besides, the reasons to think otherwise are far from convincing. It seems, therefore, that we should remain skeptical about the philosophical prospects of compatibilism and incompatibilism.

References

Abelson, Robert (1968): "Knowledge and Belief." In: *Journal of Philosophy* 65(22), pp. 725–732.
Black, Carolyn (1971): "Knowledge without Belief." In: *Analysis* 31(5), pp. 152–156.
Buckwalter, Wesley/Rose, David/Turri, John (2015): "Belief Through Thick and Thin." In: *Noûs* 49 (4), pp. 748–775.
DeRose, Keith (1992): "Contextualism and Knowledge Attributions." In: *Philosophy and Phenomenological Research* 52, pp. 913–929.
Feit, Neil/Cullison, Andrew (2011): "When Does Falsehood Preclude Knowledge?" In: *Pacific Philosophical Quarterly* 92, pp. 283–304.
Gettier, Edmund (1963): "Is Justified True Belief Knowledge?" In: *Analysis* 24, pp. 121–123.
Hemp, David (ONLINE): "The KK (Knowing That One Knows) Principle." In: *Internet Encyclopedia of Philosophy*. <http://www.iep.utm.edu/kk-princ/>
Hintikka, Jaakko (1962): *Knowledge and Belief. An Introduction to the Logic of Two Notions*. Ithaca. New York: Cornell University Press.
Hintikka, Jaakko. (1970): "'Knowing That One Knows' Reviewed." In: *Synthese* 21, pp. 141–162.

Hocutt, Max (1972): "Is Epistemic Logic Possible?" In: *Notre Dame Journal of Formal Logic* 13(4), pp. 433–453.

Lehrer, Keith (1990): *Theory of Knowledge*. Boulder: Westview Press.

Lewis, David (1996): "Elusive Knowledge." In: *Australasian Journal of Philosophy* 77, pp. 549–567.

Mannison, Don (1976): "'Inexplicable Knowledge' Does Not Require Belief." In: *Philosophical Quarterly,* 26(103), pp. 139–148.

Myers-Schulz, Blake/Schwitzgebel, Eric (2013): "Knowing That p Without Believing That p." In: *Noûs* 47(2), pp. 371–384.

Nozick, Robert (1981): *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Rescher, Nicholas (2005): *Epistemic Logic. A Survey of the Logic of Knowledge*. Pittsburgh: University of Pittsburgh Press.

Radford, Colin (1966): "*Knowledge – by Examples*." In: *Analysis* 27(1), pp. 1–11.

Rose, David/Schaffer, Jonnathan (2013): "Knowledge Entails Dispositional Belief." In: *Philosophical Studies* 166, pp. 19–50.

Russell, Bertrand (1992): *The Analysis of Mind* (11th ed.). London: Routledge.

Tokarz, Marek (1990): *"*On the Logic of Conscious Belief." *Studia Logica* 49(3), pp. 321–332.

Williamson, Timothy (2000): *Knowledge and Its Limits*. New York: Oxford University Press.

Wittgenstein, Ludwig. (2009): *Philosophical Investigations* (Revised 4th ed.). Malden, MA/Oxford: Wiley-Blackwell.

Index of Subjects

Index of Names