DE GRUYTER
MOUTON

Savas L. Tsohatzidis

TRUTH, FORCE, AND KNOWLEDGE IN LANGUAGE

ESSAYS ON SEMANTIC AND PRAGMATIC TOPICS

TRENDS IN LINGUISTICS

Savas L. Tsohatzidis
Truth, Force, and Knowledge in Language

Trends in Linguistics Studies and Monographs

Volume 344

Savas L. Tsohatzidis

Truth, Force, and Knowledge in Language

Essays on Semantic and Pragmatic Topics

DE GRUYTER
MOUTON

For Olga and Sophia

Acknowledgements

I am grateful to Daniel Van Olmen, editor of the Trends in Linguistics series, to Volker Gast, former editor of the Trends in Linguistics series, and to Barbara Karlson and Julie Miess, project editors at De Gruyter, for their support at various stages of the preparation of this book.

All but one of the essays comprising the book are based on papers that I have published in various journals and thematic volumes over the years; most of these appear here in reworked versions.

Chapter 1 is based on "Truth Ascriptions, Falsity Ascriptions, and the Paratactic Analysis of Indirect Discourse", *Logique et Analyse*, Vol. 58, No. 232 (2015), pp. 527–534.

Chapter 2 draws on two earlier papers: its first section is based on "The Hybrid Theory of Mixed Quotation", *Mind*, Vol. 107, No. 427 (1998), pp. 661–664; its second section is based on "Lost Hopes and Mixed Quotes", *Belgian Journal of Linguistics*, Vol. 17 (2003), pp. 213–229.

Chapter 3 is based on "Self-Reference and the Divorce between Meaning and Truth", *Logic and Logical Philosophy*, Vol. 22 (2013) pp. 445–452.

Chapter 4 is based on "Performativity and the 'True/False Fetish'", in Savas L. Tsohatzidis (ed.), *Interpreting J. L. Austin: Critical Essays* (Cambridge: Cambridge University Press, 2018), pp. 96–118.

Chapter 5 draws on two earlier papers: its main text is based on "Speaking of Truth-Telling: The View from *Wh*-complements", *Journal of Pragmatics*, Vol. 19 (1993), pp. 271–279; its appendix is based on "More Telling Examples: A Response to Holton", *Journal of Pragmatics*, Vol. 28 (1997), pp. 629–636.

Chapter 6 is a reprint of "The Distance Between 'Here' and 'Where I Am'", *Journal of Philosophical Research*, Vol. 40 (2015), pp. 13–21.

Chapter 7 is a reprint of "A Problem for a Logic of 'Because'", *Journal of Applied Non-Classical Logics*, Vol. 25 (2015), pp. 46–49.

Chapter 8 is based on "What 'Lack' Needs to Have: A Study in the Semantics of Privation", in John R. Taylor and Robert E. MacLaury (eds.), *Language and the Cognitive Construal of the World* (Berlin and New York: De Gruyter, 1995), pp. 81–93.

Chapter 9 draws on two earlier papers: its main text is based on "A Fake Typicality Constraint on Asymmetric Acceptability", *Cognitive Linguistics*, Vols. 3–4 (1992), pp. 391–398; its appendix is based on "Fake Means Fake: A Response to Cruse's Reply", *Cognitive Linguistics*, Vols. 3–4 (1992), pp. 403–408.

Contents

Part III **Knowledge matters** — 297

Introduction

The idea of a natural language whose expressions no one could ever understand is evidently self-contradictory. But it is far from evident what understanding an expression of a natural language (that is, knowing what is meant by that expression) consists in. Philosophers of language have been proposing and disputing answers to that question for a long time, usually taking limited amounts of linguistic data as sufficient for illustrating their views; linguists, for their part, have typically taken one or another of the philosophers' answers for granted, and have sought to extend their application, under the banners of either 'semantics' or 'pragmatics', to linguistic data of ever increasing variety and complexity. There are several things about natural language understanding that have been learned in the process, and several things that remain to be learned. It is not unreasonable to inquire, however, whether there might also be things in this complicated area that one would have to, so to speak, 'unlearn'—either because specific analyses of particular instances of linguistic understanding would have been shown to be misguided despite their initial plausibility, or because widely held theses about what linguistic understanding is and how it should in general be analysed would have been shown to have implications that are empirically or conceptually inadequate, despite being presupposed by large sections of relevant research. The essays comprising this book raise questions that lead to the recognition of some problems of these two kinds, and they thus aim to contribute to a clearer understanding of some descriptive and foundational issues that arise at the interface between linguistic semantics and pragmatics, on the one hand, and the philosophy of language, on the other. Before summarising the contributions of the individual essays, I will briefly describe some main themes under which their various topics fall.

The book's first part, "Matters of Meaning and Truth", comprises essays most of which critically examine implications or implementations of the idea, which lies at the core of many linguistic and philosophical approaches to semantics, that understanding a declarative sentence of a natural language, and therefore knowing what is meant by it, is knowing what its truth conditions are—that is, knowing which states of the world would be such that, if they obtained, the sentence would be true, and if they did not obtain, the sentence would be false. The essays in this part are of three kinds. Some of them directly dispute the idea that a truth-conditional theory can be an adequate theory of declarative sentence meaning, by arguing that some of its assumptions or implications are unsustain-

able (by arguing, for example, that there exist truth-evaluable declarative sentences that have exactly the same meanings even though they cannot have the same truth conditions; or by arguing that there are good reasons to accept, and no good reasons to reject, the view that certain declarative sentences are fully meaningful even though they do not have any truth conditions). Some other essays examine whether the truth-conditional conception of declarative sentence meaning can consistently deal with problems whose solution is widely acknowledged to be an essential condition on its viability (whether, for example, the so-called 'paratactic' analysis of indirect discourse can be maintained—and can thus obviate the problems besetting standard truth-theoretic treatments of referential opacity—without abandoning fundamental tenets of the truth-conditional conception of declarative sentence meaning that it was aiming to represent as defensible in the first place). Finally, some other essays, without questioning the truth-conditional conception of declarative sentence meaning as such, dispute particular applications, or particular invocations, that have been made of it, and specific conclusions that have been drawn on the basis of such applications or such invocations (they dispute, for example, whether the alleged truth-conditional non-equivalence of certain indexical adverbs with certain indexical adverbial phrases justifies drawing a general distinction between the so-called 'assertoric content' and the so-called 'ingredient sense' of a declarative sentence; or whether the alleged truth-conditional equivalence of sentences where an entity is described as lacking a certain property and sentences where the same entity is described as not having that property constitutes an argument in favour of a so-called 'cognitivist', and against a so-called 'objectivist', conception of declarative sentence meaning.)

The book's second part, "Matters of Meaning and Force", comprises essays most of which critically examine implications or implementations of the idea, which is central to many linguistic and philosophical approaches to pragmatics, that understanding a sentence (declarative or non-declarative) of a natural language, and therefore knowing what is meant by it, requires knowing certain conditions of the sentence's use—in particular, that it requires knowing, on the one hand, under what conditions it could be used to perform what illocutionary acts and, on the other hand, under what conditions it could be used to convey what conversational implicatures. The essays in this part are of three kinds. Some of them dispute key assumptions on which standard use-conditional accounts of linguistic interpretation have been developed (by arguing, for example, there is no viable notion of propositional content that is invariant across illocutionary act types, and that therefore an illocutionary act cannot in general be represented— as is standardly assumed—as the result of attaching an illocutionary force to a

force-impermeable propositional content; or by arguing that there are no necessary correlations between types of mental states and types of illocutionary acts, and that therefore linguistic intentionality cannot be analysed—as is standardly assumed—as a kind of projection of mental intentionality). Some other essays question whether standard use-conditional accounts of linguistic interpretation can consistently account for phenomena that have been supposed to be optimally accountable in their terms (they argue, for example, that there exist types of conversational implicature such that their derivation by the standard schema of conversational implicature derivation would require the incoherent assumption that a speaker simultaneously contravenes at least one conversational maxim and does not contravene any conversational maxim; and they also argue that one of the many sorts of things that speakers can conversationally implicate is that their hearers are conversationally uncooperative, which is a kind of implicature that the standard account of conversational implicature derivation cannot explain by reference to its fundamental assumption that speakers and hearers are conversationally cooperative, unless it abandons its even more fundamental assumption that speakers and hearers are rational agents). Finally, some other essays, without questioning the key assumptions on which standard use-conditional accounts of linguistic interpretation operate, dispute particular applications that have been made of those assumptions, and particular claims that have been based on those applications (for example, the claim that the analysis of the illocutionary act of promising enables the rejection of the widely shared view that evaluative statements cannot be logically derived from wholly non-evaluative ones; or the claim that illocutionary acts of declaration provide the logically necessary scaffolding for explaining how institutional facts emerge out of non-institutional facts).

The book's third part, "Knowledge Matters", comprises three essays that, like the essays of the first two parts, discuss either questions of meaning in relation to truth or questions of meaning in relation to force; their specific focus in discussing such questions, however, is the interpretation of certain types of knowledge attributions and of knowledge denials; and their results are consequently relevant not only to semantic or pragmatic issues related to the philosophy of language, but also to certain epistemological issues (for example, the issue of whether or not truth is a necessary condition of knowledge; the issue of whether or not the concept of assertion is definable in terms of the concept of knowledge; and the issue of whether linguistic knowledge should be analysed in dispositionalist, action-invoking, or in anti-dispositionalist terms, not action-invoking, terms).

There are, of course, hundreds of problems and issues other than the few investigated in this book that one would have to examine in detail before attempting with any plausibility to advance general theses about the nature of natural language understanding, and of the respective roles of semantics and of pragmatics in its elucidation. Thus, even though the essays collected here indicate perspectives from which, and conditions under which, one might seek to improve one's understanding of the specific problems that they discuss, they certainly do not purport to be promoting an all-purpose recipe for the conduct of either semantic or pragmatic investigation. The principal thing I would hope them to achieve, if they are in the right direction, is to prompt readers to rethink some issues that are generally supposed to have already been settled, or that might appear easy to settle, and to draw the appropriate methodological morals from the realisation that even seemingly settled or seemingly easy issues are often not what they seem to be. I now proceed to some brief indications on the contents of the individual essays.

Part I. Matters of meaning and truth

Chapter 1 ("Truth ascriptions, falsity ascriptions, and the paratactic analysis of indirect discourse") offers a critical assessment of Donald Davidson's influential 'paratactic' analysis of the logical form of indirect discourse, and explores the metasemantic implications of the results of that assessment. I first present certain obviously valid types of inference involving indirect speech reports as premises and truth or falsity ascriptions as conclusions, and show that the validity of these types of inference is incompatible with Davidson's paratactic analysis of the logical form of indirect discourse. I then argue that this particular family of counterexamples to the Davidsonian analysis of indirect discourse has a deeper significance for Davidson's overall project of using what he terms a 'Tarski-style' truth theory as a theory of natural language meaning: that project assumes that grasp of the concepts of truth and falsity is essential to natural language interpretation, yet Davidson's analysis cannot characterise as valid certain natural language inferences whose recognition as valid is arguably constitutive of one's grasp of the concepts of truth and falsity. Finally, I argue that Davidson's analysis not only has implications that are in tension with his proposed justification for the kind of semantic program that he recommends, but, in addition, would require for its defence the explicit adoption of assumptions that are characteristic of a kind of semantic program that he rejects.

Chapter 2 ("The hybrid theory of mixed quotation") is a critical examination of Ernie Lepore's and Herman Cappelen's proposal to analyse the logical form of

sentences involving mixed quotation in terms of a 'sametokening' relation borrowed from Davidson's analysis of direct quotation and a 'samesaying' relation borrowed from Davidson's analysis of indirect quotation. The chapter shows that the proposed analysis fails to specify sufficient conditions for mixed quotation and, more importantly, that it does not succeed in specifying necessary conditions either, for three reasons: first, because sentences employing mixed quotation can be true in situations in which the sametokening clause of their proposed Davidsonian analysis is not satisfied (even though the samesaying clause might be); secondly, because sentences employing mixed quotation can be true in situations in which the samesaying clause of their proposed Davidsonian analysis is not satisfied (even though the sametokening clause might be); and finally, because sentences employing mixed quotation can be true even in situations in which neither the sametokening nor the samesaying clause of their proposed Davidsonian analysis is satisfied.

Chapter 3 ("Self-reference and the divorce between meaning and truth") argues that the widely held thesis that a declarative sentence's meaning is identical to its truth condition faces not only the well-known problem that there exist pairs of declarative sentences whose members have different meanings even though they have necessarily identical truth conditions, but also the hitherto unnoticed, and converse, problem that there exist pairs of declarative sentences whose members have identical meanings even though they have necessarily different truth conditions. The chapter further argues that this new problem—which is caused by a certain type of self-referential sentence—arises both at the inter-linguistic and at the intra-linguistic level, and that it cannot be assimilated to certain superficially similar problems that the thesis under discussion could independently be argued to be vulnerable to.

Chapter 4 ("Performativity and the 'true/false fetish'") examines the contemporary semantic significance of J. L. Austin's thesis that explicit performative utterances are not truth-evaluable despite their declarative grammatical form—a thesis that Austin regarded as obviously true and not in need of argumentation, but which, if true, would enable a potent objection to the truth-conditional conception of linguistic content, since admitting both it and the truth-conditional conception of content would require embracing the obviously false conclusion that explicit performatives utterances are contentless. I begin by observing that the nearly universal rejection of Austin's thesis by contemporary philosophers of language is supposed to be justified by an abductive argument to the effect that unless one assumed, contra Austin, that explicit performatives are truth-evaluable, one could not *explain* what everyone, including Austin, would regard as their most distinctive and remarkable feature, namely that, in issuing them,

speakers can accomplish the illocutionary acts that they thereby name. I then argue that attempts to justify the denial of Austin's non-truth-evaluability thesis by producing explanations of performativity that essentially depend on the hypothesis that explicit performatives are truth-evaluable cannot succeed for at least two types of reason: on the one hand, because utterances that, on the proposed explanations, should be capable of being explicit performative ones turn out to be incapable of being explicit performative ones; on the other hand, because utterances that, on the proposed explanations, should be incapable of being explicit performative ones turn out to *be* capable of being explicit performative ones. Since the source of both of these explanatory failures turns out to be none other than the adoption of the hypothesis that explicit performative are truth-evaluable, I suggest that they strongly undermine the anti-Austinian view and vindicate Austin's thesis, in favour of which I then sketch an independent argument based on the behaviour of explicit performatives in deductive inferential contexts—specifically, on the fact that their behaviour in such contexts could not be reconciled with the hypothesis that they are truth-evaluable unless one denied the applicability, in those contexts, of certain logically fundamental inference rules. I conclude that the Austinian thesis that explicit performatives are not truth-evaluable can by no means be regarded as having been superseded, and that, consequently, the threat it poses to the truth-conditional conception of declarative sentential content remains fully in force.

Chapter 5 ("Speaking of truth-telling: The view from *wh*-complements") critically examines the widely accepted thesis, originating with Zeno Vendler in philosophy and with Lauri Karttunen in linguistics, that sentences reporting the contents of an act of telling by means of a *wh*-complement entail that the subject of the act of telling has been truthful (in contrast to sentences reporting the contents of an act of telling by means of a *that*-complement, which do not entail that the subject of the act of telling has been truthful). The chapter argues (a) that the thesis in question is demonstrably false; (b) that its falsity has certain non-negligible implications on certain wider linguistic and philosophical issues; (c) that a thesis that would be crucially different from it in that it would replace reference to entailments of truthfulness by reference to defeasible presumptions of truthfulness would be both accurate and capable of wider application (in particular, that it would apply both to *wh*-complements and to *that*-complements of "tell"); and (d) that the reasons behind the truth of that *other* thesis could be sought in a theory of linguistic conventions as solutions to coordination problems between information-seeking individuals.

Chapter 6 ("The distance between 'here' and 'where I am'") scrutinises Michael Dummett's claim that consideration of the semantic behaviour of the adverb "here" and of the adverbial phrase "where I am" justifies drawing a fundamental distinction (whose neglect, in Dummett's view, has been the source of important philosophical errors) between two features of the meaning of a declarative sentence which he proposes to call, respectively, 'assertoric content' and 'ingredient sense'. The chapter offers a variety of linguistic arguments showing that the evidence to which Dummett appeals in order to support his proposed distinction does not, in fact, support it, and that, consequently, the philosophical arguments he has advanced on its basis are not compelling.

Chapter 7 ("A problem for a logic of 'because'") examines a logical system designed with the purpose of elucidating the meaning of the natural language connective "because", in those of its uses that exhibit a non-causal explanatory sense. The chapter shows that the system in question is unsuccessful relative to its self-assigned goal, since its introduction-rules for "because" have the undesirable consequence of, on the one hand, not validating inferences that ought to be validated, and, on the other hand, validating inferences that ought not to be validated. The chapter further points out that both of these defects arise from the system's failure to distinguish between cases where "because" governs explicit truth ascriptions to sentences from cases where it governs those sentences themselves, and argues that the sensitivity of "because" to that crucial distinction would be predictable on the independently defensible assumption that it is a hyperintensional connective, i.e. a connective in the context of which inter-substitution of logically equivalent sentences fails to preserve truth-value.

Chapter 8 ("What 'lack' needs to have: A study in the semantics of privation") takes as its point of departure an argument of George Lakoff's according to which the verb "lack" compels us to realise that lexical meaning is not a matter of relationships holding between words and mind-independent worlds—a view that Lakoff identifies as the central tenet of what he calls 'objectivist' semantics—, but rather a matter of relationships holding between words and certain types of mental models of the world—a view that he regards, and recommends, as the central insight of what he calls 'cognitive' semantics. The chapter has two aims. The first is to show that the semantic analysis of "lack" that Lakoff offers is seriously defective and hence not really serviceable in the context of the 'objectivism' vs. 'cognitivism' dispute. The second is to sketch an alternative analysis of the semantics of "lack", which, though far removed from what Lakoff would regard as 'objectivist' assumptions, is equally far removed from the kind of 'cognitivism' that Lakoff is interested in promoting.

Chapter 9 ("A fake typicality constraint on asymmetric acceptability") disputes an argument to the effect that inclusion of typicality traits in lexical semantic description is justified by the fact that, unless such traits were assumed to be part of lexical meaning, one could not explain the asymmetric acceptability of compound sentences of the forms "S but S′" and "S but ~S′", where S is a sentence in which a property p is predicated of an entity t and S′ is a sentence in which a different property q is predicated of the same entity t. The chapter contends that this argument, which is endorsed by several theorists of lexical semantics, fails to confer semantic significance on typicality traits, for at least three reasons: first, because the acceptability pattern exhibited by substitution instances of the two sentential frames can be asymmetric even in the absence of relevant typicality contrasts; secondly, because the acceptability pattern exhibited by substitution instances of the two sentential frames can be symmetric even in the presence of relevant typicality contrasts; and finally, because the very assumption that, in the cases under consideration, judgments of acceptability in the linguistic domain are causally related to judgments of typicality in the extra-linguistic domain is rendered doubtful by the fact that judgements of the former kind would be possible for subjects for whom judgments of the latter kind would be impossible.

Chapter 10 ("Correlative and noncorrelative conjunctions in argument and nonargument positions") makes a preliminary contribution to the investigation of the question as to what the semantic difference is between coordinate structures of the form "X and Y" and coordinate structures of the form "both X and Y", by critically examining an interesting semantic proposal that lies at the core of a recent syntactic analysis of the relation between structures of the two types, when they occupy argument positions. I argue that the semantic proposal in question fails to properly differentiate between the two structures (and that the syntactic claims it was meant to support are consequently undermined). I also argue, however, that there is clear independent evidence that the two structures *are* semantically distinct, at least when truth-conditionally relevant differences are taken to be semantically relevant differences; and that the evidence in question points to the need for a cross-categorial account of that semantic difference, since it is a difference ascertainable both when the structures occupy argument positions and when they occupy nonargument positions.

Part II. Matters of meaning and force

Chapter 11 ("Yes–no questions and the myth of content invariance") argues that, despite their apparent simplicity and obvious centrality to linguistic communication, yes–no questions cannot be adequately handled by John Searle's highly influential theory of speech acts—specifically, that neither the system of speech act

analysis nor the system of speech act classification that that theory proposes can be consistently applied to them. The chapter further argues that both the analytical and the classificatory problems that Searle's theory faces in dealing with yes–no questions originate from its adherence to the fundamental assumption, inspired by Frege, that the contents to which the various types of illocutionary force get attached are invariably propositions. Finally, the chapter argues that rejecting this 'content invariance' thesis, and allowing different types of illocutionary force to determine different types of content as their attachment sites, opens up the way for alternative schemes of speech act analysis and speech act classification, in the context of which the problems besetting Searle's theory, and several other theories that it has influenced, fail to arise.

Chapter 12 ("Deontic trouble in speech act botany") considers Searle's claim that, by adopting the one or the other of two definitions of permission that he provides (and which are assumed to be equivalent, given his definition of prohibition), one could resolve the problem of accommodating permissions into his category of directive illocutionary acts, despite the fact that permissions (unlike, most obviously, orders) appear not to satisfy what he takes to be the defining feature of the category of directive illocutionary acts. The chapter advances three theses: first, that even if the proposed two definitions of permission were correct in themselves, they could not possibly resolve Searle's classification problem; second, that they are not, in fact, correct in themselves; and third, that they are not equivalent to each other (because the definition of prohibition that underlies the assumption of their equivalence is itself incorrect). On the basis of this evidence, the chapter concludes that if, as Searle and many others assume, a reasonable condition of adequacy on a theory of speech acts is that it correctly analyse, and, in so doing, reflect the natural affinities between the concepts of ordering, prohibiting and permitting, Searle's theory of speech acts fails to satisfy that condition.

Chapter 13 ("The gap between speech acts and mental states") develops a critique of Searle's foundational thesis that the intentionality of speech acts is not intrinsic to them but derives from the intrinsic and underived intentionality of mental states. I begin by noting that Searle's main linguistic argument for that thesis resides in the claim that unless one assumed that illocutionary acts of all major types are necessarily expressions of mental states of corresponding major types (for example, unless one assumed that all assertive acts are necessarily expressions of beliefs, all directive acts are necessarily expressions of desires, all commissive acts are necessarily expressions of intentions, etc.), one could not explain the alleged fact that it is always unacceptable to conjoin the attempted performance of an illocutionary act with the explicit denial of the possession of

its corresponding mental state (for example, that it is always unacceptable to attempt to assert something while denying that one believes it, to attempt to request that something be done while denying that one wants it to be done, to attempt to promise to do something while denying that one intends to do it, etc.) I then argue that Searle's claim is open to two lines of objection: the first consists in showing that the linguistic facts he appeals to do not *require* the kind of explanation he proposes, and that an alternative explanation, which does not entitle him to the claim that there are necessary connections between kinds of speech acts and kinds of mental states, can be made available; the second consists in showing that there are other relevant linguistic facts that *exclude* the kind of explanation he proposes, and that can only be understood by assuming that there exist *no* necessary connections between kinds of speech acts and kinds of mental states. On the basis of this evidence, I conclude that Searle's project of deriving the intentionality of language from the intentionality of mind is less well motivated than it is usually taken to be.

Chapter 14 ("A purported refutation of some theories of assertion") defends normative analyses of assertion against the claim, due to Peter Pagin, that there is a 'method of refutation' capable of showing that every normative analysis of assertion is wrong. I first show that, even if the proposed 'method of refutation' were supposed to be reliable, it could not be used, in the way it was meant to be used, either against the kinds of theories of assertion that Pagin was aiming to undermine or in favour of the kind of theory of assertion that he was aiming to support. I then show that the proposed 'method of refutation' could not be supposed to *be* reliable, since its operation would presuppose the denial of fundamental distinctions that any adequate analysis of assertion should respect.

Chapter 15 ("Two consequences of hinting") has two aims. The first is to defend, against known arguments to the contrary, the claim that hinting constitutes a counterexample to Searle's so-called 'principle of expressibility', i.e. to the thesis that every illocutionary act that is performed inexplicitly can in principle be performed explicitly. The second is to argue that certain instances of hinting provide counterexamples, of a type that has so far not been recognised, to Searle's descriptivist analysis of proper names, i.e. to the thesis that each proper name is semantically equivalent to a cluster of identifying descriptions of the individual to which it refers.

Chapter 16 ("How to test a test for perlocutionary act names") examines the most sophisticated attempt since Austin to provide a grammatical test for separating perlocutionary from non-perlocutionary act verbs. The chapter argues that the proposed test is unsuccessful on three counts: first, it marks as perlocutionary

act verbs certain verbs that are not action verbs at all; second, it marks as perlocutionary act verbs two types of action verbs that are clearly not perlocutionary act verbs; and third, it marks as non-perlocutionary act verbs certain action verbs that are clearly perlocutionary act verbs. The chapter concludes by suggesting that, rather than searching for a grammatical mark of the perlocutionary/non-perlocutionary divide, one should maintain a purely semantic characterisation of that divide, and that apparent obstacles to maintaining such a semantic characterisation can be overcome by recognising that several speech act denoting verbs participate in a pattern of regular polysemy whereby they are systematically ambiguous between an illocutionary and a perlocutionary interpretation.

Chapter 17 ("Speaker meaning, sentence meaning, and metaphor") examines a key claim of Searle's account of metaphor, namely, the claim that metaphorical meaning is a variety of speaker meaning rather than of sentence meaning. The chapter argues that this claim is problematic in two fundamental respects. The first problem is that even if one assumes, as Searle requires, that metaphorical meaning is a variety of speaker meaning rather than of sentence meaning, Searle's theoretical proposals as to how metaphorical meaning should be distinguished from other acknowledged varieties of speaker meaning unavoidably lead to contradictory statements when applied to relevant data. The second problem is that Searle's assumption that metaphors *are* functions of what speakers, as opposed to sentences, mean, is inconsistent with the results of applying to metaphorical utterances the test that is most widely accepted, and that Searle himself independently accepts, as affording a reasoned decision as to whether a given conveyed meaning is or is not a function of what a sentence, as opposed to a speaker, means, since that test unequivocally pronounces metaphorical meanings to be functions of what sentences mean, rather than functions of what speakers of sentences choose to mean by them.

Chapter 18 ("Voices and noises in the theory of speech acts") is an extended review essay evaluating two collective volumes dedicated to linguistic and philosophical research in speech act theory. Besides offering critical assessments of the individual contributions to those volumes, it includes general discussion of broader empirical, conceptual, and methodological issues in that field, as well as some reflections on aspects of its historical development.

Chapter 19 ("Searle's derivation of promissory obligation") defends the thesis, originating with Hume, that no evaluative conclusions are logically derivable from wholly non-evaluative premises against a famous argument by Searle that purports to demonstrate its falsity by claiming that statements about promises entail statements about obligations (specifically, that statements about what an individual has promised to do entail statements about what that individual ought

to do). The chapter develops two main arguments. The first is that, even if the entailment relation posited by Searle were supposed to hold, it would not amount to a demonstration of the falsity of the Humean thesis, since Searle's required supplementary assumption that statements about promises, unlike statements about obligations, are non-evaluative is open to serious doubt, and arguably in conflict with Searle's own independent work on the normative foundations of linguistic institutions. The second argument is that, independently of how the distinction between evaluativity and non-evaluativity is to be applied to statements about promises or to statements about obligations, there are cogent logico-linguistic reasons against Searle's central claim that statements about promises *entail* statements about obligations, the most significant of those reasons being one that derives from the fact that statements about promises are semantically non-extensional in a specific sense in which statements about obligations are not. The chapter concludes that Searle's argument does not succeed in undermining the Humean thesis and notes the implications of that fact on certain broader aspects of Searle's work.

Chapter 20 ("Searle's *Making the Social World*") scrutinises the central thesis of Searle's book, *Making the Social World*, according to which all institutional facts are created by speech acts of declaration. I argue that Searle's articulation and defence of that thesis is problematic in three fundamental respects. First, in order to avoid the infinite regress that would ensue if his central thesis were held conjointly with his independent claim that linguistic facts are institutional facts, Searle is forced the radically reduce the scope of the central thesis—by transforming it into the much narrower thesis that all *non*-linguistic institutional facts are created by speech acts declaration—and thus violates his own methodological principle that there should be a *single* kind of explanation for the creation of *every* kind of institutional fact. Second, even restricted to non-linguistic institutional facts, the thesis that all such facts are created by speech acts of declaration relies on illicitly assimilating the few cases where a declaration creates an institutional fact with the many cases where it does not create an institutional fact but only creates a *condition* for the creation of an institutional fact; and when the confusion underlying that assimilation is cleared up, it transpires that, at best, only a relatively small *subset* of non-linguistic institutional facts can be held to be literally created by declaration. Finally, even as an account of the creation of that small subset of non-linguistic institutional facts, Searle's account is rendered problematic by his parallel requirement that an institutional fact cannot exist unless and until it is collectively *accepted* as existing, since that requirement is in tension with the thesis that such a fact owes its existence solely to an individual's

act of declaration, which is defined by Searle as an act whose performance is by itself *sufficient* for bringing into existence the fact that it represents as existing.

Chapter 21 ("A paradox of cooperation in the theory of implicatures") identifies a type of conversational implicature that cannot be calculated by a hearer using the Gricean schema of conversational implicature calculation unless the hearer assumes that the speaker can be cooperative only if she is taking the hearer to be un-cooperative. The existence of conversional implicatures of this kind, I argue, implies that the Gricean theory of conversational implicatures faces the dilemma of having to accept either that the calculation of conversational implicatures violates general principles of rationality or that it does not rely on a presumption of mutual cooperation between conversational participants.

Chapter 22 ("An inferential impasse in the theory of implicatures") shows that the Gricean explanation of two prominent types of putative conversational implicatures faces a so far unnoticed problem when confronted with utterances that simultaneously carry implicatures of *both* of these prominent types. The problem, in a nutshell, is that, since the Gricean theory requires implicatures of these two types to be calculated under mutually incompatible inferential regimes, it cannot without inconsistency derive implicatures of either type when a single utterance carries both of them. After explaining how this problem arises, the chapter indicates why it would fail to arise if certain distinctively anti-Gricean, but independently supported, assumptions about utterance interpretation were adopted.

Part III. Knowledge matters

Chapter 23 ("How to forget that 'know' is factive") critically examines a study by Allan Hazlett that, on the one hand, alleges that the acceptability of a certain type of knowledge attribution requires the wholesale rejection of the thesis, which is virtually universally adopted in epistemology, that knowledge is truth-entailing (i.e. that "know", when introducing a *that*-complement, is semantically factive) and, on the other hand, proposes that the truth-commitment communicated by knowledge attributions is simply a conversational implicature of those attributions. The chapter argues that Hazlett's position ought to be rejected for a variety of reasons, three of the most important of which are the following. (a) The position is methodologically unstable since it depends on interpreting the acceptability of certain instances of a conjunctive schema as evidence of the absence of contradictoriness, while refusing to interpret the *un*acceptability of certain other instances of the *same* conjunctive schema as evidence of the presence of contradictoriness. (b) The position is incapable of giving any account of independently

available linguistic evidence which unequivocally indicates that "know" is semantically ambiguous between a factive, truth-entailing, sense, and a non-factive, non-truth-entailing, sense. (c) The position's recourse to the hypothesis that the truth-commitment communicated by knowledge attributions is simply a conversational implicature of those attributions is unsuccessful, since that hypothesis is inconsistent with every reliable procedure of either identifying or deriving conversational implicatures, and in particular with the requirement that nothing could be a conversational implicature unless it were capable of being cancelled.

Chapter 24 ("Three problems for the knowledge rule of assertion") scrutinises a key argument that Timothy Williamson has used in motivating the so-called 'Knowledge First' program in contemporary epistemology. Williamson's argument is that unless the speech act of assertion were supposed to be governed by the so-called Knowledge Rule—the rule that one should assert only what one knows—, it could not be explained why sentences of the form "*A* and I do not know that *A*" are unassertable even though they might be true if asserted. The chapter advances three arguments against Williamson's proposed explanation of the unassertability of sentences of the form "*A* and I do not know that *A*". The first two of these arguments show that, even assuming that Williamson's explanandum has been properly circumscribed, his explanation would be incorrect, and the third shows that his explanandum has not been properly circumscribed.

Chapter 25 ("Grammars as objects of knowledge: The availability of dispositionalism") concerns the opposition between the anti-dispositionalist interpretation of grammatical knowledge, which maintains that such knowledge exists whether or not it can be behaviourally manifested, and the dispositionalist interpretation, which identifies grammatical knowledge with the in principle possibility of certain types of behavioural manifestation. The chapter makes two complementary contributions to the debate between dispositionalist and anti-dispositionalist interpretations of grammatical knowledge: on the one hand, it refutes a prominent argument, due to Chomsky, in favour of the anti-dispositionalist interpretation; on the other hand, it presents an original argument, inspired by Wittgenstein, in favour of the dispositionalist interpretation. Like the broader debate to which they belong, both contributions are conceptual in character, and make essential reference to, on the one hand, logically crucial similarities between ascriptions of knowledge and ascriptions of ability, and, on the other hand, logically crucial differences between statements ascribing abilities and statements ascribing abilities to exercise abilities. The contributions thus exemplify one way in which attention to the semantics of knowledge attributions and ability attributions can shed light on issues in the epistemology of language.

Part I   **Matters of meaning and truth**

Chapter 1
Truth ascriptions, falsity ascriptions, and the paratactic analysis of indirect discourse

1 Introduction

The present essay considers certain obviously valid types of inferences involving indirect speech reports as premises and truth or falsity ascriptions as conclusions, and argues that their validity is incompatible with Davidson's so-called "paratactic" analysis of the logical form of indirect discourse (Davidson 1968). It further argues that this particular failure of the Davidsonian analysis has a special significance for Davidson's overall project of using what he terms a "Tarski-style" truth theory as a theory of natural language meaning (Davidson 1967): that project assumes that grasp of the concepts of truth and falsity is essential to natural language interpretation, yet Davidson's analysis cannot characterise as valid certain natural language inferences whose recognition as valid is arguably constitutive of one's grasp of the concepts of truth and falsity. Finally, the essay argues that Davidson's analysis not only has implications that are in tension with his proposed justification for the kind of semantic program that he recommends, but, in addition, would require for its defence the explicit adoption of assumptions that are characteristic of a kind of semantic program that he rejects.

2 Indirect speech reports and Davidsonian semantics

When A is not self-referential, inferences of the following forms are evidently valid:

(i)   S said that A, A ∴ S said something true
(ii)   S said that A, $\sim A$ ∴ S said something false
(iii)   S said that $\sim A$, $\sim A$ ∴ S said something true
(iv)   S said that $\sim A$, A ∴ S said something false

For example, in each of the following arguments, the conclusion obviously follows from the premises:

(1)   Galileo said that the Earth moves.
         The Earth moves.
         Therefore, Galileo said something true.

(2)   Galileo said that the Earth moves.
         The Earth does not move.
         Therefore, Galileo said something false.

(3)   Galileo said that the Earth does not move.
         The Earth does not move.
         Therefore, Galileo said something true.

(4)   Galileo said that the Earth does not move.
         The Earth moves.
         Therefore, Galileo said something false.

Not only are such inferences obviously valid, but a person's ability to recognise their validity can plausibly be held to be constitutive of that person's possession of the concepts of truth and falsity: one would not be credited with understanding what truth and falsity are if one was unable to acknowledge that the conclusions of arguments such those in (1)–(4) do indeed follow from their premises. And it is presumably for that reason that Aristotle implicitly appeals to inferences of this sort in order to *define* truth and falsity, in the famous passage of *Metaphysics* that many (including Tarski 1944: 342–343) have regarded as the beginning of sense in inquiries about the concept of truth:

> To *say* of what is that it is not, or of what is not that it is, is false, while to *say* of what is that it is, or of what is not that it is not, is true (*Metaphysics* Γ. 7, 1011b26–27; emphasis added)[1]

It is appropriate, then, to demand of an adequate semantic account of indirect speech reports, and especially of an account that, like Davidson's, purports to be a truth-theoretic one, to be in a position to acknowledge the validity of inferences

---

1 Translation by W. D. Ross, in Barnes (1984).

of this kind. What I propose to argue now, however, is that Davidson's account is not in that position.

As is well known, on Davidson's analysis of indirect speech reports the utterer U of a sentence of the form "S said that P" accomplishes, logically speaking, exactly two things: First, U demonstratively refers by the singular term "that" to an utterance of a sentence that has the same form as P but is *not* part of the indirect speech report (logically speaking, that is, the sentence "S said that P" ends just after the word "that", and that word is a demonstrative pronoun referring to an independently occurring utterance rather than a complementiser introducing a constituent clause). And secondly, U states that there has been an utterance by S that has the same content as (and so, in Davidsonian parlance, that 'samesays') the utterance to which U demonstratively refers by using the singular term "that". Davidson provides the following succinct statement of his analysis, as applied to the indirect speech report "Galileo said that the Earth moves":

> The paratactic semantic approach to indirect discourse tells us to view an utterance of 'Galileo said that the Earth moves' as consisting of the utterance of two sentences, 'Galileo said that' and 'The Earth moves'. The 'that' refers to the second utterance, and the first utterance is true if and only if an utterance of Galileo's was the same in content as ('translates') the utterance to which the 'that' refers. (Davidson 1976: 39/1984: 176–177)

For present purposes, there are two points that it is important to keep in mind in considering Davidson's analysis. The first is that, since, according to that analysis, what *appear* to be subordinate clauses in indirect speech reports are not, in reality, syntactic or semantic parts of those reports at all, but are simply the real word *objects* to which the demonstrative singular terms allegedly occurring in the reports refer, these clauses cannot be contributed by the reports to any *inferences* in which the reports occur as premises. To suppose that an argument beginning with the premise "Galileo said that.", where "that" demonstratively refers to an exhibited linguistic object, can properly include that object itself as one of its further premises is no more coherent than to suppose that an argument beginning with the premise "Galileo liked that.", where "that" demonstratively refers to an exhibited painting, can properly include that painting itself as one of its further premises. This means that, on any coherent reconstruction of the arguments in (1)–(4) along Davidsonian lines, the seemingly subordinate clauses of the first premises should *not* be regarded as being transformable into additional independent premises (a transformation that, even if it were not incoherent, would

have the unwelcome consequence of producing arguments with redundant premises in the case of (1) and (3), and arguments with inconsistent premises in the case (2) and (4)), but should simply be *excised* from the arguments on the grounds that maintaining them as additional independent premises would amount to confusing vehicles of representation with objects of representation. Thus, the only parts of the first premises of (1)–(4) that would survive in a coherent reconstruction of those arguments along Davidsonian lines would be the parts ending with the demonstratively construed "that"; and each one of those parts should be interpreted in the way stipulated by Davidson, i.e. as a statement to the effect that a linguistic object produced by Galileo in the past has the same content as the linguistic object that is being demonstratively referred to by the utterer of "that" in the present.

The second point that it is important to keep in mind in considering Davidson's analysis is that, just as saying, of two objects to which one refers, that they have the same colour, is not saying *what* their colour is, so saying, of two utterances to which one refers, that they have the same content, is not saying *what* their content is. If, for example, you know neither Finnish nor Turkish, and someone tells you that a certain Finnish utterance to which he is pointing with his left hand has the same content as a certain Turkish utterance to which he is pointing with his right hand, he has not told you (no matter how sincere towards you he might be and how knowledgeable about Finnish and Turkish he might be) *what* the content of either the Finnish utterance or the Turkish utterance is. In the same way, saying, as Davidson's analysis does, that, in order for an indirect speech report to be true, the utterance object that is thereby attributed to the reported speaker must have the same content as the utterance object that is being demonstratively referred to by the reporting speaker, is not saying *what* the content of either of these utterance objects is.

To see, now, that the Davidsonian analysis is not in a position to account for the obvious validity of inferences such as (1)–(4), it is sufficient to observe what these inferences would amount to, if the indirect speech reports they contain were interpreted in conformity with the Davidsonian analysis. Let Donald be the person who is uttering the demonstratives that, on the Davidsonian analysis, the inferences contain. Then, from the viewpoint of the Davidsonian analysis, the inference in (1) is tantamount to an inference (call it (D-1)) in which, from the premises that the Earth moves, and that Galileo has produced an utterance object that has the same content as the utterance object that is currently being demonstratively referred to by Donald, it is concluded that Galileo said something true. And the inference in (2) is tantamount to an inference (call it (D-2)) in which, from the

premises that the Earth does not move, and that Galileo has produced an utter-ance object that has the same content as the utterance object that is currently being demonstratively referred to by Donald, it is concluded that Galileo said something false. (Analogous reformulations would be easy to provide for the re-maining two cases.)

It is evident, however, that inferences such as (D-1) or (D-2) are not valid: From the statement that the Earth moves (or that the Earth does not move), and that Galileo has produced an utterance object that is identical in content to the utterance object that is currently being demonstratively referred to by Donald, *nothing whatsoever* follows about the truth (or falsity) of anything said by Galileo. Of course, something might *be* capable of being shown to follow if, from the state-ment that Galileo has produced an utterance object that is identical in content to the utterance object that is currently being demonstratively referred to by Don-ald, one could deduce *what* the content of the utterance object produced by Gali-leo was. As already noted, however, from the statement that two utterances are identical in content one cannot deduce what the content of either is; and so, the statement that the utterance produced by Galileo in the past and the utterance that is the object of Donald's current demonstration have the same content no more allows one to deduce what the content of either of these utterances is than the statement that a demonstratively referred to utterance of Finnish and a demonstratively referred to utterance of Turkish are identical in content allows one to deduce what the content of either the Finnish utterance or the Turkish ut-terance is.

Davidson's analysis of indirect speech reports, therefore, cannot avoid re-garding as invalid certain obviously valid inferences involving such reports. And it cannot avoid this precisely because of what Davidson has described as its "novel" claim "upon [which] everything depends"—namely, that

> from a semantic point of view the content-sentence in indirect discourse [i.e. the sentence following 'said that' in an indirect speech report, SLT] *is not con-tained in the sentence whose truth counts*, i.e. the sentence that ends with 'that'. (Davidson 1968: 143/1984: 106; emphasis added)

For, if what follows "said that" in the first premises of the inferences in (1)–(4) is semantically *not* part of those premises, then there can be no semantic relation whatsoever between those premises *and the second premises* of (1)–(4). If, how-ever, there is no semantic relation whatsoever between the first and the second premises of (1)–(4), there is no way for the conclusions of (1)–(4) to follow from their premises, and the inferences must accordingly be held to be invalid. Given,

then, that the inferences are clearly valid, and that the least that one would ex-
pect from a satisfactory account of the logical form of indirect speech reports is
that it enable one to represent as valid all clearly valid inferences in which such
reports occur as premises, the conclusion must be that Davidson's account fails
as an account of the logical form of indirect speech reports.

3 Indirect speech reports and Davidsonian metasemantics

Though the problem exposed above is by no means the only problem that David-
son's account faces, it has a special significance that distinguishes it from several
other problems that have been raised and discussed in the literature that the ac-
count has so far generated (for an overview of that literature, written from a per-
spective sympathetic to Davidson, see Sennet 2013). As is well known, the entire
Davidsonian program in natural language semantics rests on the assumption that
what Davidson describes as a "Tarski-style" truth theory for a natural language
could serve as a theory of meaning for that language, in the sense that a person's
knowledge of that truth theory would suffice for that person's interpreting the
utterances of the language's users.[2] It can plausibly be held, however, that a per-
son will not be able to understand what a truth theory states if it has no idea what
truth and falsity are, and that it will have no idea what truth and falsity are if it is
unable to recognise, among other things, that certain inferences involving the
truth and falsity predicates, such as those exemplified in (1)–(4), are valid. And
since, as argued above, a person who understands indirect speech reports in ac-
cordance with Davidson's "paratactic" account will *not* be able to recognise that
inferences such as (1)–(4) are valid, it follows that it will be not be a person capa-
ble of understanding what truth and falsity are and, therefore, of using a truth
theory of the sort envisaged by Davidson in order to interpret the utterances of
others. It appears, then, that inferences of the sort we have been considering are
significant not only by virtue of showing that Davidson's account of the logical
form of indirect speech reports is, considered in itself, unsuccessful, but also by

---

2 Notice that, as Davidson acknowledges (1973: 321/1984: 134; cf. 2005: 7–48), this assumption
reverses what Tarski was taking to be the proper order of explanation: The "Tarski-style" theory
envisaged by Davidson is supposed to elucidate the concept of meaning by taking the concept
of truth for granted, whereas Tarski's actual theory of truth was supposed to elucidate the con-
cept of truth by taking the concept of meaning for granted.

virtue of showing that that account is in tension with one of the fundamental assumptions of the Davidsonian semantic program.

It may finally be noted that Davidson's analysis of indirect speech reports not only has implications that, as just seen, are in tension with his proposed justification for the kind of semantic program that he recommends, but, in addition, would require for its defence the explicit adoption of assumptions that are characteristic of a kind of semantic program that he *rejects*. As is well known, Davidson denies that a theory of translation can serve as a theory of interpretation (and so, as a theory of meaning in his sense). And the main reason he offers for this denial is that

> we can know which sentences of the subject language translate which sentences of the object language without knowing what any of the sentences of either language mean (Davidson 1973: 317/1984: 129)

But this Davidsonian objection to translational theories of interpretation has fundamentally the same source as the objection that, as we saw, can and should be raised, in view of the problem posed by inferences such as those in (1)–(4), against Davidson's own analysis of indirect discourse. For, as previously argued, the obvious validity of these inferences could not be held to be consistent with what the Davidsonian analysis of indirect discourse stipulates *unless* one were to *mistakenly* suppose that stating that two expressions have the same meaning is the same thing as stating *what* their meaning is. So, the mistake that, according to Davidson, one would be making if one were to suppose that a theory of translation can serve as a theory of interpretation is precisely the mistake that one would *have* to make if one were to suppose that the validity of inferences such as (1)–(4) is consistent with his analysis of indirect discourse. It seems, therefore, that an extra reason that, in view of such inferences, Davidson would have for abandoning the "paratactic" analysis is that abandoning it would allow him to maintain his principal argument in favour of the thesis that a theory of translation cannot serve as a theory of interpretation.

The fact that Davidson has not, apparently, realised that his "paratactic" analysis of indirect discourse cannot be held conjointly with his objection to translational theories of interpretation may of course be due simply to the fact that he didn't happen to consider the role of indirect speech reports in inferences such as those in (1)–(4). For it is the role of indirect speech reports in such inferences that makes especially clear what might otherwise not be apparent, namely, that it is one thing to say, as the "paratactic" analysis does, that a certain demon-

strated utterance is "the same in content as ('translates')" a certain attributed utterance (Davidson 1979: 39/1984: 177) and quite another to say *what* the content of either of these utterances is. It may also be, however, that Davidson was, at least intermittently, under the confused impression that saying the former of these things *is* the same as saying the latter, and was thus led to the mistake of supposing that his analysis does provide what it ought to provide but does not in fact provide. Thus, at one point in which he purports to be informally explicating the "paratactic" analysis, Davidson says that, on that analysis, "what follows [the demonstrative 'that'] gives the content of the subject's saying" (Davidson 1968: 142/1984: 106). As we have seen, however, what the "paratactic" analysis actually claims is *not* that the reporting speaker's demonstrative "that" refers to an utterance that 'gives the content' of an utterance attributed to the reported speaker; rather, what it claims is *only* that the reporting speaker's demonstrative "that" refers to an utterance that is "the same in content as ('translates')" an utterance attributed to the reported speaker. And these two claims can be held to equivalent only by someone who commits the error that, according to Davidson himself, proponents of translational theories of interpretation would be committing—in other words, only by someone who fails to realise that saying of two utterances that they have the same content is *not* the same thing as 'giving the content' of either (i.e. is not the same thing as saying *what* the content of either is). I suspect that this is not the only case where what a Davidsonian analysis actually offers is not the same thing as what it is advertised by Davidson as offering, and that an examination of similar cases in other places of Davidson's oeuvre might be instructive.

Chapter 2
The hybrid theory of mixed quotation

1 Two problems for the hybrid theory

In their "Varieties of Quotation", Herman Cappelen and Ernie Lepore (1997a) point out that the phenomenon of mixed quotation is much more widespread than the amount of semantic attention so far paid to it would lead one to suspect, argue that some prominent theories of either direct or indirect quotation cannot successfully be adapted in order to account for it, and propose to successfully account for it by combining the demonstrative account of indirect quotation due to Davidson (1968) with a demonstrative account of direct quotation inspired by Davidson (1979).

Direct quotation is exemplified by (1), and its Davidsonian analysis is assumed to be as in (2), where "ST" stands for the sametokening relation (a relation that holds between two expressions just in case they have the same shape):

(1)   Alice said "Life is difficult to understand".
(2) $\exists u$(Says (a, u) & ST(u, these)). Life is difficult to understand.

Indirect quotation is exemplified by (3), and its Davidsonian analysis is assumed to be as in (4), where "SS" stands for the samesaying relation (a relation that holds between two expressions just in case they have the same content):

(3)   Alice said that life is difficult to understand.
(4) $\exists u$(Says(a, u) & SS(u, that)). Life is difficult to understand.

Mixed quotation is exemplified by (5), and its correct analysis should, according to Cappelen and Lepore, combine the distinctive feature of the Davidsonian analysis of direct quotation with the distinctive feature of the Davidsonian analysis of indirect quotation. Specifically, its correct analysis should be as in (6), where a demonstratively referred to utterance is claimed to stand, by virtue of its content, in the samesaying relation, and, by virtue of the shape the one or more of its constituent expressions, in the sametokening relation, to an utterance produced by the quoted subject:

(5)   Alice said that life "is difficult to understand".

(6) $\exists u$(Says(a, u) & SS(u, that) & ST(u, these)). Life is difficult to understand.

We may call this proposal "the hybrid theory of mixed quotation".

Unfortunately, the hybrid theory fails to provide a satisfactory analysis of mixed quotation. This is because both the sametokening and the samesaying relations that it invokes force it to give incorrect truth conditions to several types of sentences involving mixed quotation.

To see the undesirable consequences of invoking the sametokening relation, consider the following sentences:

(7)   Descartes said that man "is a thinking substance".
(8)   Frege said that predicate expressions "are unsaturated".
(9)   Socrates said that an unexamined life "is not worth living for a human".

These sentences are capable of supplying accurate information about certain views that Descartes, Frege and Socrates have expressed, and it is hard to imagine how they could do that unless they could be true. Their truth, however, hardly requires that Descartes, Frege and Socrates used *English* in order to express their views, and it is indeed possible to combine them, without semantic oddity, with adverbial modifiers where this possibility is explicitly excluded, as in the sentences below:

(10)   In one of the greatest philosophy books ever written in Latin, Descartes said that man "is a thinking substance".
(11)   In a German text that is often cited by modem semanticists, Frege said that predicate expressions "are unsaturated".
(12)   In a famous Greek dialogue, Socrates said that an unexamined life "is not worth living for a human".

On the hybrid theory, however, the respective equivalents of (7), (8) and (9) would be the following:

(13) $\exists u$(Says(d, u) & SS(u, that) & ST(u, these)). Man is a thinking substance.
(14) $\exists u$(Says(f, u) & SS(u, that) & ST(u, these)). Predicate expressions are unsaturated.
(15) $\exists u$(Says(s, u) & SS(u, that) & ST(u, these)). An unexamined life is not worth living for a human.

These could certainly not be true unless Descartes, Frege and Socrates had used *English* in order to express their views, since their truth would require Descartes, Frege and Socrates to have *sametokened*, at least in part, the *English* sentences whose tokens are thereby demonstrated. The sametokening relation invoked by the hybrid theory cannot, therefore, be legitimately utilised, since it imputes to sentences involving mixed quotation a truth condition that they do not in fact have.

To see the undesirable consequences of invoking the samesaying relation, consider the following sentences:

(16)   Alice said that this five meters long bar "is ten meters long".
(17)   Alice said that the Queen of England "is the Queen of England's mother".
(18)   Alice said that these very healthy people "are in very bad health".

In order for these sentences to be true, which they can very well be, it is certainly *not* necessary that Alice should have uttered any obvious contradictions. In particular, it is not necessary that she should have uttered sentences expressing the view that a five meters long bar is a ten meters long bar, or the view that the Queen of England is the Queen of England's mother, or the view that a very healthy person is a person in very bad health. Indeed, any speaker of (16), (17) and (18) can, without semantic oddity, append to these sentences explicit denials of the view that Alice did express any such contradictory views, as in (among others) the sentences below:

(19)   Alice said that this five meters long bar "is ten meters long"—without, of course, realising, let alone claiming, that it is five meters long.
(20)   Alice said that the Queen of England "is the Queen of England's mother"—without, of course, realising, let alone claiming, that she is the Queen of England.
(21)   Alice said that these very healthy people "are in very bad health"—without, of course, realising, let alone claiming, that they are very healthy.

On the hybrid theory, however, the respective equivalents of (16),(17) and (18) would be the following:

(22) $\exists u(\text{Says}(a, u) \& \text{SS}(u, \text{that}) \& \text{ST}(u, \text{these}))$. This five meters long bar is ten meters long.
(23) $\exists u(\text{Says}(a, u) \& \text{SS}(u, \text{that}) \& \text{ST}(u, \text{these}))$. The Queen of England is the Queen of England's mother.

(24) $\exists u(\text{Says}(a, u) \,\&\, \text{SS}(u, \text{that}) \,\&\, \text{ST}(u, \text{these}))$. These very healthy people
         are in very bad health.

These could certainly not be true unless Alice had expressed the obvious contradictions therein demonstrated, since their truth would require her to have *samesaid* these obvious contradictions—that is to say, to have uttered sentences that are identical in content with them. The samesaying relation invoked by the hybrid theory cannot, therefore, be legitimately utilised, since it imputes to sentences involving mixed quotation a truth condition that they do not in fact possess.

Since neither the samesaying nor the sametokening clause introduced by the hybrid theory can be maintained, it seems that Davidsonian accounts of either direct or indirect quotation are no more capable of jointly providing a satisfactory analysis of mixed quotation than the other prominent accounts of direct and indirect quotation reviewed by Cappelen and Lepore.

2 The full case against the hybrid theory[1]

Preliminaries

I have argued so far that hybrid theory of mixed quotation propounded by Cappelen and Lepore (hereafter, C&L) fails to specify necessary conditions on mixed quotation, for two reasons: first, because sentences employing mixed quotation can be true in situations in which the sametokening clause of their proposed Davidsonian analysis would not be satisfied (even though the samesaying clause might be); and secondly, because sentences employing mixed quotation can be true in situations in which the samesaying clause of their proposed Davidsonian analysis would not be satisfied (even though the sametokening clause might be). In what follows I will complete my case against the hybrid theory by (a) showing that C&L's subsequent attempt to neutralise these two arguments—which I will respectively call "the sametokening argument" and "the samesaying argument"—has been unsuccessful and (b) presenting two additional arguments against C&L's analysis, the first of which shows that sentences involving mixed

---

1 The paper on which the first part of the present essay is based (Tsohatzidis 1998) has given rise to a reply by Cappelen and Lepore (Cappelen and Lepore 1998), to which I responded in Tsohatzidis (2003). The second part of the present essay is based on that response.

quotation can be true even in situations in which the samesaying and the sametokening clauses of their Davidsonian analysis would *simultaneously* fail to be satisfied, and the second of which complements the previous ones by questioning not the alleged necessity of the samesaying and the sametokening clauses, but rather their alleged sufficiency.

The sametokening problem

The sametokening argument against the hybrid theory, it will be recalled, was that sentences like (7) and (8), repeated below for convenience,

(7)   Descartes said that man "is a thinking substance".
(8)   Frege said that predicate expressions "are unsaturated".

cannot be supposed to have, as the hybrid theory proposes, the logical forms in (13) and (14), respectively,

(13) $\exists u(\text{Says}(\text{d}, u) \;\&\; \text{SS}(u, \text{that}) \;\&\; \text{ST}(u, \text{these}))$. Man is a thinking substance.
(14) $\exists u(\text{Says}(\text{f}, u) \;\&\; \text{SS}(u, \text{that}) \;\&\; \text{ST}(u, \text{these}))$. Predicate expressions are unsaturated.

since (13) and (14) have entailments that (7) and (8) do not have. Specifically, the truth of (7) and (8) does *not* require that Descartes or Frege have used, in order to express the views that (7) and (8) truthfully ascribe to them, any expressions that are tokens of the same expression-types that the *English* expressions quoted in (7) and (8) are tokens of; by contrast, the truth of (13) and (14) *does* require that Descartes and Frege have used, in order to express those views, expressions that are tokens of the same expression-types that those *English* expressions are tokens of.

In a short paper purporting to reply to my sametokening and samesaying arguments (Cappelen & Lepore 1998), C&L characterise the examples to which the sametokening argument appeals as "mixed quotes in which the direct quoted part is a translation of the words used in the quoted utterances" and declare themselves "prejudiced in favour of a unified account" of mixed quotation that would encompass both these "translational" mixed quotes and the "standard" mixed quotes studied in their original paper (all quotations from Cappelen & Lepore 1998: 665; notice that no distinction between 'standard' and 'non-standard' mixed quotes was employed anywhere in C&L's original paper, and that no analysis of that distinction is attempted in the text of their reply). At this point, one might expect C&L either to actually propose the "unified account" of mixed quo-

tation in favour of which they say they are predisposed (and which, when available, would supposedly cover both "translational" and "standard" mixed quotes) or to postpone the construction of such a "unified account" to a future occasion. Astonishingly, however, C&L do neither of these things, suggesting instead that their original account just *is* the sought for "unified account". But how could this be, given that their account, by virtue of the sametokening clause it contains, makes obviously incorrect predictions when applied to mixed quotes of the "translational" variety? C&L's *entire* answer to that crucial question is that they place "no a priori constraints on what can same-token what" and that, in particular, they "see no principled objection to saying that in some contexts same-tokening can rely on translation" (quotations from Cappelen & Lepore 1998: 665).

That answer is entirely unsatisfactory, though seeing that it is may require a few more words than it itself contains. The first part of the answer, in which C&L present themselves as jettisoning all "a priori constraints on what can sametoken what", is probably not meant to be taken seriously, and cannot, in any case, be so taken, since *some* a priori constraints would obviously be part of *any* account. For example, even C&L would presumably require that sametokening is not a relation that *everything* bears to *everything* else; but the requirement that sametokening is not a relation instantiated by just any pair of objects is certainly an a priori, rather than an a posteriori, constraint on sametokening. Similarly, even C&L would presumably require that, since the sametokening relation is a particular kind of similarity relation, it must be symmetric rather than non-symmetric; but the requirement that the sametokening relation should, *qua* similarity relation, be symmetric rather than non-symmetric is certainly an a priori, rather than a posteriori, constraint on sametokening, deriving as it does from an a priori, rather than a posteriori, constraint on the notion of similarity. It would seem, then, that C&L could not really mean what they say when they deny commitment to *any* a priori constraints. But since they don't say which a priori constraints they like and which they don't, it is best to leave that matter to one side and turn to the remaining, and less insubstantial, part of their answer, which is to the effect that their theory can avoid the counterexamples in (7) and (8) by imposing the (presumably, a posteriori) constraint that "in some contexts same-tokening can rely on translation".

The first thing to notice about this 'constraint' is that, whether or not it achieves its purpose, it involves a radical, and unacknowledged, change in C&L's conception of the sametokening relation. The notion of sametokening originally employed by C&L was meant to be a refined version of the central idea of Davidson's analysis of direct quotation—the idea, namely, that, in direct quotation, what the quoting speaker does is, on the one hand, to demonstratively refer to an

utterance and, on the other hand, to claim that the quoted speaker produced an utterance *with the same shape* as the utterance demonstratively referred to (notice that, in his reaction to a recycled version of C&L's analysis, Davidson considers the, according to him, "unimportant" question whether the refinements suggested by C&L would really be required, and explicitly denies that they would, pointing out that the thought that they would relies on a misunderstanding, on C&L's part, of his conception of the notion of shape; see Davidson 1999: 101). Now, if the notion of sametokening is to be construed, as C&L were originally supposing, along these Davidsonian lines—namely, as a relation in which two expressions stand just when they have *the same shape*—, then it is clear that (7) and (8) are inescapable counterexamples to C&L's analysis of mixed quotation. For, as we have seen, the truth of (7) and (8) is consistent with Descartes' and Frege's having used only Latin and only German expressions, respectively, in order to formulate their views. And since there are no Latin and no German expressions with the same content *and the same shape* as the English expressions quoted in (7) and (8), it follows that the truth of (7) and (8) is consistent with neither Descartes' nor Frege's having ever *sametokened* those English expressions, contrary to what C&L's analysis predicts. If the original understanding of the sametokening relation remains in force, then, C&L's analysis is as clearly mistaken as an analysis could be.

C&L's reaction to this situation has the merit of not retreating to the false claim that the truth of (7) and (8) does require Descartes and Frege to have used *English* in order to express their views. It does not, however, have the merit of explicitly acknowledging that their proposed way of dealing with the situation involves not a reformulation but a *rejection* of their original understanding of the sametokening relation. For, their proposed way of dealing with the situation consists in saying that "same-tokening can rely on translation"—in other words, that Descartes and Frege may well have used only Latin and only German expressions, respectively, in order to formulate the views truthfully reported in (7) and (8), but that this can be made consistent with the idea that their expressions *sametoken* the English expressions quoted in (7) and (8), *if* one assumes that expressions belonging to two different languages are to be counted as sametokening each other as long as they are *translations* of each other. That assumption, however, whether or not it leads to an acceptable new theory (about which more below), amounts to a complete rejection of the original understanding of the sametokening relation: On anyone's account, expressions from two different languages that are *translations* of each other are expressions that have the *same content* even though they may have completely *different shapes*. Given, then, that translation is a process that aims at content-preservation *rather than* shape-

preservation, and given, on the other hand, that it is *shape*-preservation rather than *content*-preservation that was the defining characteristic of the original account of the sametokening relation deriving from Davidson, C&L's reaction to the problem posed by (7) and (8) involves using a notion of sametokening that is not merely different from but actually *opposite* to the notion of sametokening with which they have been working. Instead, therefore, of misleadingly suggesting that their original account of mixed quotation can deal with (7) and (8) *as it stands*, they should simply acknowledge that (7) and (8) compel them to put forward a *new* account of mixed quotation, and that the particular account that they have chosen to put forward (whether as a substitute for or as a supplement to the old one) introduces a notion of sametokening that is the exact opposite of the notion they have been using.

The next question is whether the new account is less unsuccessful than the old one. Unfortunately, it is just as unsuccessful as the old one, though for different reasons. Let us assume, as C&L now propose, that sametokening is a relation that holds between two expressions when they are 'translations' of each other—that is, when their *content* is the same, even though their *shapes* may be wildly different. Of course, anyone expecting that a definition of the sametokening relation should be consistent with generally held assumptions about the distinction between expression types and expression tokens would find the new definition unacceptable, since, interpreted as a way of recapturing the type/token distinction, it has obviously absurd implications (it implies, for example, that a token of the English word "day" and a token of the French word "jour" are tokens of the *same* word, simply because they happen to have the same content, or that a token of the word "pianoforte" and a token of the word "fortepiano" are tokens of the *same* word, simply because they both happen to refer to the same musical instrument). Nevertheless, since C&L would have probably disowned by now any commitment to mean by the term 'sametokening' what everyone else would expect that term to mean (this is, perhaps, what they were really gesturing at by expressing their dislike of all 'a priori constraints' on sametokening), it would be unfair to fault their new account on these grounds. The real problem with the new account is that, by proposing an understanding of sametokening based on the notion of translation—and so, exclusively on *content*-preservation, rather than on *shape*-preservation—the new account makes the sametokening relation indistinguishable from the *samesaying* relation, contrary to what C&L's general views about quotation explicitly require.

Recall that, on C&L's original exposition of their views, direct quotation, indirect quotation and mixed quotation were *different* phenomena requiring *distinct* accounts, and that the accounts proposed by C&L were aiming to respect

that distinctness requirement by analysing direct quotation exclusively in terms of the notion of sametokening, indirect quotation exclusively in terms of the notion of samesaying, and mixed quotation in terms of a combination of the notions of sametokening and of samesaying. Clearly, the distinctness requirement could not be coherently fulfilled if sametokening and samesaying turned out to be *the very same thing*. But if, as C&L now propose, sametokening is a relation that holds between two utterances when they are 'translations' of each other—that is, when they have the same *content*, irrespective of their possible differences in *shape*—, then sametokening *cannot* be distinct from samesaying. For, the notion of samesaying that C&L claimed they were assuming is the notion of samesaying introduced in Davidson's (1968) analysis of indirect quotation. And the notion of samesaying introduced in Davidson's analysis of indirect quotation just *is* the notion of two utterances' having the same *content* (and, being, in that sense, 'translations' of each other) irrespective of their possible differences in shape. As Davidson quite explicitly puts it at one point, the claim that, in indirect quotation, the quoting speaker asserts that a demonstratively indicated utterance *samesays* an utterance produced by the quoted speaker, is equivalent to the claim that, in indirect quotation, the quoting speaker asserts that a demonstratively indicated utterance has "the same content as ('translates')" an utterance produced by the quoted speaker (Davidson 1976: 39). If, therefore, the new sametokening relation proposed by C&L is what they say it is—namely, a relation holding between two utterances when they are translations of each other—, then the notion of sametokening simply *is* the very same notion as the notion of samesaying they were presenting themselves as assuming. But—and this is the crucial point—if the notions of sametokening and samesaying are the very same notion, then their own analytical requirements are massively and irreparably violated: not only must their analysis of *mixed* quotation be now recognised to have been multiply mistaken (first because either its samesaying or its sametokening clause becomes redundant, and secondly because, whichever of these clauses is dropped, the analysis fails to properly distinguish mixed quotation from either direct quotation or indirect quotation), but the very possibility of distinguishing *direct* from *indirect* quotation itself evaporates (if sametokening and samesaying are the very same thing, it is impossible for an analysis of direct quotation in terms of sametokening to be saying anything distinct from an analysis of indirect quotation in terms of samesaying).

I conclude that C&L's attempt to avoid the implications of the sametokening argument I presented against their analysis of mixed quotation puts them in a position noticeably worse than the position they originally occupied. Had they stood by their initial conception of the sametokening relation (which they had

derived from Davidson), they would have to concede that their analysis of mixed quotation is incorrect, but they could at least retain the analyses of direct and of indirect quotation they are inheriting from Davidson. By changing their conception of sametokening beyond all recognition, however, they succeed in arriving at a situation where, contrary to their own requirements, they can properly account neither for direct quotation, nor for indirect quotation, nor for mixed quotation.

The samesaying problem

The samesaying argument against the hybrid theory, it will be recalled, was that sentences like (16) and (17), repeated below for convenience,

(16)   Alice said that this five meters long bar "is ten meters long".
(17)   Alice said that the Queen of England "is the Queen of England's mother".

cannot be supposed to have, as the hybrid theory proposes, the logical forms in (22) and (23), respectively,

(22) $\exists u(\text{Says}(a, u) \,\&\, \text{SS}(u, \text{that}) \,\&\, \text{ST}(u, \text{these}))$. This five meters long bar is ten meters long.
(23) $\exists u(\text{Says}(a, u) \,\&\, \text{SS}(u, \text{that}) \,\&\, \text{ST}(u, \text{these}))$. The Queen of England is the Queen of England's mother.

since (22) and (23) have entailments that (16) and (17) do not have. Specifically, the truth of (16) or of (17) does *not* require Alice to have expressed the self-contradictory view that a certain five meters long bar is a ten meters long bar, or the self-contradictory view that the Queen of England is the Queen of England's mother; by contrast, the truth of (22) and of (23) *does* require Alice to have expressed precisely those self-contradictory views, since it requires Alice to have produced utterances that *samesay*—that is, that have the same content as—the patently self-contradictory sentences "This five meters long bar is ten meters long" and "The Queen of England is the Queen of England's mother".

In the short paper purporting to reply to my sametokening and samesaying arguments, C&L simply say that the samesaying argument relies on a "misreading" of their original paper. The argument, they note, "assumes that we think that the samesay relation holds between all and only utterances that are 'identical in content'". But, they point out, "We don't." (All quotations from Cappelen & Lepore 1998: 665.) Since this is *all* that C&L say in reply, someone might wonder why they didn't take the trouble to even briefly enlighten their readers as to what

the conception of samesaying assumed in their original paper really *is*, if it is not a conception based on sameness of content. The answer to that question is that they couldn't possibly provide such enlightenment, since the conception of samesaying assumed in their original paper simply *is* the conception according to which two utterances samesay each other just in case they have the same content. More specifically, the *only* thing that C&L's original paper says about samesaying is that its authors rely on the understanding of that notion provided in Davidson's account of indirect quotation (indeed, C&L make it a point to emphasise that they are merely interested in showing how Davidson's account of indirect quotation, *assuming it to be correct*, can be "exploited" in their own account of mixed quotation, but disavow any intention "to engage in an evaluation" of it; see Cappelen & Lepore 1997a: 442). However, Davidson's account of indirect quotation explicitly defines the samesaying relation as a relation holding between utterances just in case they have the same content (as we have already seen, utterances samesaying each other are, according to Davidson, all and only utterances that have "the same content" (Davidson 1976: 39), or, as he also puts it in the paper that C&L cite, that "match in content" (Davidson 1968: 145)). Consequently, anyone reading C&L's original paper and wishing to suppose that the notion of samesaying they employ is not *meaningless* has no other option but to suppose that it has the meaning that Davidson stipulated for it. In which case, of course, employing Davidson's interpretation of samesaying in reading their original paper not only does not constitute a "misreading" of that paper but rather constitutes the only course of action capable of avoiding the hypothesis that the paper employs the notion of samesaying without giving it *any* meaning.

Of course, that particular interpretation of C&L's paper might be taken to have been unduly charitable, if C&L's real position on the notion of samesaying turned out to be that that notion cannot, in fact, be given *any* substantive meaning. And, unfortunately, this is what their real position appears to be, if one is to judge from their reply. For their reply cites, purportedly as a clue to their real position on the samesaying relation, a different paper of theirs published after the one that was the target of my criticisms (Cappelen & Lepore 1997b), in which, along with repudiating certain theses about the goals of semantics held, according to them, both by Davidson and by virtually every other contemporary philosopher of language, they explicitly *abandon* Davidson's account of indirect quotation on the grounds that it imposes "a priori constraints on what can samesay what" (Cappelen & Lepore 1997b: 291) and avoids "closely attending to our actual practice of indirect reporting" (Cappelen & Lepore 1997b: 293). Now, the various claims of that later paper—none of which had been used in the earlier one—have been extensively and effectively criticised in the literature (see Reimer 1998 and,

especially, Richard 1998), but these criticisms need not be invoked in the present connection. What does need to be noticed is simply that, even if all of C&L's later anti-Davidsonian claims *were* correct, no alternative definition of the samesaying relation would be derivable from them. For, obviously, delivering sermons against 'a priori constraints on samesaying' and thereby rejecting Davidson's account of indirect quotation does not amount to actually proposing any *positive* account of indirect quotation or of samesaying. And, in the later paper, C&L make it perfectly clear that they simply have *no* positive account to offer: from their criticisms of Davidson's account of indirect quotation, they tell us, "no *positive* theory of indirect quotation follows" (Cappelen & Lepore 1997b: 291; italics original). But if no positive theory of indirect quotation follows, no positive characterisation of samesaying can follow either; indeed, the only characterisation of samesaying that, in the context of C&L's later paper, becomes possible is the following completely vacuous one: samesaying is that relation between reported utterances and reporting clauses which is such that, *if* we knew what it is—which we *don't*—, we would know when a sentence containing an indirect quotation is true and when it is false. But this is not an *alternative* to Davidson's or to anyone else's account of samesaying or of indirect quotation. It is just a refusal to commit oneself to any substantive account. And the refusal to undertake substantive commitments can hardly be construed as a way of fulfilling substantive commitments: if reference to the samesaying relation really gives, as C&L's original paper was claiming, the truth conditions of sentences containing indirect quotations, then the samesaying relation could not non-vacuously be defined, as C&L's later paper suggests, as the relation which is such that we don't know what it is, but that, if we knew what it is, we would know when a sentence containing an indirect quotation is true and when it is false. It seems, then, that, instead of protesting that they have been "misread", C&L should simply acknowledge that they are facing a dilemma: either to rely on Davidson's definition of samesaying, and draw the negative implications that that reliance turns out to have on their analysis of mixed quotation, or to claim that no substantive definition of samesaying can possibly be given, in which case, of course, they would have been left with no substantive analysis to defend.

My conclusion is that C&L's reaction to the samesaying argument I presented against their hybrid theory of mixed quotation makes things worse for them than they already were. Had they remained faithful to the content-based understanding of samesaying that a charitable interpretation of their original text demands, they could at least retain the account of indirect quotation that they inherit from Davidson, though they would obviously have to abandon their account of mixed

quotation (since, in view of the proposed counterexamples, retaining that account would require endorsing the preposterous claim that contradictory and non-contradictory sentences have the same content). By effectively refusing to give *any* interpretation to the samesaying clauses that their analyses contain, however, they remain with no real analysis of *either* mixed *or* indirect quotation (and, a fortiori, with no means of differentiating either of these phenomena from the phenomenon of direct quotation).

The combined problem

The two arguments against the hybrid theory of mixed quotation presented so far have relied on instances of mixed question which show that the one or the other, but not both, of the samesaying and samesaying conditions invoked by that theory can fail to be satisfied. We are now going to see that there also exist instances of mixed quotation which show that the samesaying and samesaying conditions invoked by the theory can *simultaneously* fail to be satisfied, and which thus provide independent evidence against the theory's fundamental claim that these conditions are necessary features of every instance of mixed quotation.

Consider the following dialogues between John and his sister Jane, where Jane appeals to her mother's authority in order to dispute the correctness of her brother's use of certain referring expressions:

(25) *John*. Let's drink this marvellous apple juice.
   Jane. Mom said that "this marvellous apple juice" is whiskey.
(26) *John*. Would you like me to play that nice violin for you?
   Jane. Mom said that "that nice violin" is a cello.

Given C&L's hybrid theory, the logical forms of Jane's utterances in these dialogues should be taken to be, respectively, the following:

(27) $\exists u(\text{Said}(m, u) \& \text{SS}(u, \text{that}) \& \text{ST}(u, \text{these}))$. This marvellous apple juice is whiskey.
(28) $\exists u(\text{Said}(m, u) \& \text{SS}(u, \text{that}) \& \text{ST}(u, \text{these}))$. That nice violin is a cello.

Now, it is clear that taking (27) and (28) to be the correct logical forms of Jane's contributions to the dialogues in (25) and (26) requires accepting that Jane's utterances in (25) and (26) could not be true unless it was Jane's *mother* who had both samesaid and at least partly sametokened the utterances demonstratively referred to in (27) and (28). It is equally clear, however, that the truth of Jane's utterances in (25) and (26) does *not* require her mother to have either samesaid or

even partly sametokened the utterances demonstratively referred to in (27) and (28). Specifically, although Jane's utterances in (25) and (26) may be meant to suggest that *her brother* has inadvertently committed himself to the obviously false views that apple juice is whiskey and that violins are cellos, the truth of Jane's utterances certainly does not require that *her mother* has committed herself to any such views, since it does not require that *her mother* has produced utterances identical in content with the obviously false utterances "This marvellous apple juice is whiskey" and "That nice violin is a cello". Indeed, Jane could without any semantic oddity expand her utterances in (25) and (26) by explicitly denying that *her mother* has ever produced utterances identical in content with those obviously false utterances. And this suffices for showing that it is not a condition on the truth of Jane's claims in (25) and (26) that *her mother* has samesaid those obviously false utterances. But neither does the truth of Jane's utterances in (25) and (26) require *her mother* to have produced any tokens of the expression "this marvellous apple juice" (as, according to Jane, *her brother* did while attempting to refer to what was, in fact, whiskey), or any tokens of the expression "that nice violin" (as, according to Jane, *her brother* did while attempting to refer to what was, in fact, a cello). Indeed, Jane could without any semantic oddity expand her utterances in (25) and (26) by stating what is anyway contextually obvious, namely, that it is only *her brother's* mistaken use of these referring expressions that she meant to be quoting while speaking about her mother. And this suffices for showing that it is not a condition on the truth of Jane's utterances in (25) and (26) that *her mother* has sametokened these referring expressions. But if the truth of Jane's utterances in (25) and (26) does not require Jane's *mother* to have produced any utterances that either samesay or sametoken the utterances demonstratively referred to in (27) and (28), it follows that Jane's utterance's in (25) and (26) are instances of mixed quotation with respect to which the samesaying and the sametokening clauses of C&L's analysis are *simultaneously* not satisfied. And these instances, of course, provide powerful additional evidence against the hybrid theory's claim that these clauses specify necessary features of every instance of mixed quotation.

It is important to notice that nothing in the above argument excludes the possibility that there might be contexts, *different* from the one suggested by the dialogues in (25) and (25), relative to which the truth of Jane's utterances *would* require her mother to have both samesaid and at least partly sametokened the utterances demonstratively referred to in (27) and (28)—one such context, for example, could (given certain assumptions) be the context where Jane responds to John exactly as she does above, but where John's utterance is neither the one figuring in (25) nor the one figuring in (26) but rather the utterance, "Tell me, Jane,

have you noticed anything recently that has made you think that mom is becoming crazy?" That possibility, however, far from comforting C&L's analysis, would reveal a deep flaw in its orientation. For, the fact that seemingly unambiguous utterances such as Jane's cannot have truth conditions that are *stable* across these and many other conversational environments suggests that, in each instance of mixed quotation, the question *who is being quoted* in that instance cannot have a *context-invariant* answer. And since the truth conditions for mixed quotation that C&L propose assume that, in each instance of mixed quotation, that question *does* have a context-invariant answer, their analysis can justifiably be taken to be no less misguided, and no more secure against potential counterexamples, than an analysis that would attempt to attribute to each sentence containing an indexical a context-invariant truth-condition.

The problem of sufficiency

Though it should be amply clear by now that neither the samesaying nor the sametokening clauses of C&L's analysis specify necessary conditions on mixed quotation, the three arguments presented so far leave open the question whether the two clauses, together with the trivially required clause about utterance-production that precedes them, might jointly specify sufficient conditions on mixed quotation. We are now going to see that they do not jointly specify sufficient conditions, either.

To begin with a problem that, though real enough, would not be unmanageable, note that the kinds of representations C&L use do not invariably equip them with a fully unambiguous format for the expression of their claims about mixed quotation. For example, the following sentences are obviously not semantically equivalent, and one should accordingly expect an adequate theory of mixed quotation to provide non-equivalent semantic representations for them.

(29) a. Ann said that "books" cost a lot.
      b. Ann said that books "cost" a lot.
      c. Ann said that books cost "a lot".
      d. Ann said that "books cost" a lot.
      e. Ann said that books "cost a lot".

However, all these sentences would receive, given the formalism C&L employ, exactly the same semantic representation, namely, (30):

(30) $\exists u(\text{Said}(a, u) \,\&\, \text{SS}(u, \text{that}) \,\&\, \text{ST}(u, \text{these}))$. Books cost a lot.

For, what (30) says is simply that Mary produced an utterance that, on the one hand, samesays, and, on the other hand, at least partly sametokens, the utterance "Books cost a lot". And *if* this is what *any* of the sentences in (29a)–(29e) means, then it is what they *all* mean. Consequently, accepted as it stands, C&L's formalism would have the effect of representing as semantically equivalent sentences that are not, in fact, semantically equivalent.

Now, this particular problem could be easily overcome. To make the formalism more discriminating, one should enrich it with a bracketing and co-indexation system that would make it possible to unambiguously express, in each proposed representation, which demonstrating elements are supposed to refer to which demonstrated elements, and would thus associate with semantically non-equivalent sentences like those in (29a)–(29e) semantically non-equivalent proposed representations like those in (29′a)–(29′e), respectively:

(29′)   a. $\exists u(\text{Said}(a, u) \ \& \ \text{SS}(u, \text{that}) \ \& \ \text{ST}(u, \text{these}_i))$. [Books]$_i$ cost a lot.
         b. $\exists u(\text{Said}(a, u) \ \& \ \text{SS}(u, \text{that}) \ \& \ \text{ST}(u, \text{these}_i))$. Books [cost]$_i$ a lot.
         c. $\exists u(\text{Said}(a, u) \ \& \ \text{SS}(u, \text{that}) \ \& \ \text{ST}(u, \text{these}_i))$. Books cost [a lot]$_i$.
         d. $\exists u(\text{Said}(a, u) \ \& \ \text{SS}(u, \text{that}) \ \& \ \text{ST}(u, \text{these}_i))$. [Books cost]$_i$ a lot.
         e. $\exists u(\text{Said}(a, u) \ \& \ \text{SS}(u, \text{that}) \ \& \ \text{ST}(u, \text{these}_i))$. Books [cost a lot]$_i$.

There is, however, a further, and this time quite serious, problem for C&L's analysis that cannot be overcome by any such notational manoeuvres. The problem is that only *some* and not *all* arrangements of quotational devices in the complement clause of a mixed quotation are attestable, and that any theory which, like C&L's, is completely unable to predict which arrangements are attestable and which are not fails to provide sufficient conditions for mixed quotation.

For example, any adequate theory (and, of course, any theory which, like C&L's, is rich in exhortations about "closely attending to our actual practice" of quotation) should be in a position to predict that (31) and (32) are attestable instances of mixed quotation whereas (33) is not, or that (34) and (35) are attestable instances of mixed quotations whereas (36) is not:

(31)   Mary said that John is "fed up" with music.
(32)   Mary said that John is fed up "with music".
(33)   * Mary said that "John is fed" up with music.
(34)   Mary said that "whom John killed" is the big question.
(35)   Mary said that whom John killed "is the big question".
(36)   * Mary said that whom John "killed is the" big question.

C&L's theory, however, is not capable of delivering these predictions. What it falsely predicts, instead, is that *all* of the above instances of mixed quotation are equally attestable, since, as far as the theory is concerned, no restrictions of any sort are stipulated that would allow one to distinguish between permissible and impermissible arrangements of quotational devices in the complement clause of a mixed quotation (to put it in the theory's terminology, no restrictions of any sort are stipulated that would allow one to distinguish between 'parts' of the complement clause that would and 'parts' of the complement clause that would not be permissible terms of the hypothesised sametokening relation).

Clearly, this problem cannot be overcome by just equipping the theory with the bracketing and co-indexation system envisaged above; indeed, such a system would merely help re-expressing the problem more precisely—namely, as the problem of finding out why *some* but not *all* of the theoretically possible co-indexations correspond to attestable as opposed to unattestable instances of mixed quotation. It seems, then, that C&L's theory should be taken to simply fail to specify sufficient conditions for mixed quotation *unless* it could be supposed that the restriction determining which co-indexations correspond to attestable instances and which to unacceptable ones is so *obvious* that C&L did not bother to explicitly state it. And the only restriction that would be a plausible candidate for obviousness is, of course, a restriction to the effect that attestable instances are those in which the quoted material in the complement clause of a mixed quotation is a syntactic constituent, whereas unattestable instances are those in which the quoted material in the complement clause of a mixed question is not a syntactic constituent (thus, the asterisked examples above would be unattestable because the quotation marks they contain do not isolate syntactic constituents, whereas the un-asterisked ones would be attestable because the quotation marks they contain do isolate syntactic constituents). Unfortunately, however, obvious solutions are not always correct ones, and the solution just suggested is a case in point. For, there are plenty of routinely attestable instances of mixed quotation where the quoted expression in the complement clause is *not* a syntactic constituent. In the exchanges below, for example, the expressions "critics liked" and "thieves have stolen" are certainly not syntactic constituents, and yet there is nothing extraordinary about the mixed quotations employed:

(37) *A*. What did John say about the book?
   B. He said that "critics liked" it.
(38) *A*. What did John say about the box?
   B. He said that "thieves have stolen" it.

The situation, then, is one in which there appear to be both quite potent *and* quite unobvious reasons why some instances of mixed quotation are attestable and some others unattestable. And since there is nothing in C&L's analysis that could help one to even suspect what these reasons are, the analysis can certainly not be supposed to provide sufficient conditions for mixed quotation.

Conclusion

We have just seen that the samesaying and sametokening clauses of C&L's analysis, together with the trivially required clause about utterance-production that precedes them, do not jointly specify sufficient conditions for mixed quotation; and we had already seen that, for at least three different types of reason, the samesaying and sametokening clauses do not specify individually necessary conditions either. Combining these two results, it would not, I suppose, be unfair to conclude that C&L's proposed contribution to the analysis of the phenomenon of mixed quotation amounts to a mere reminder that the phenomenon exists.

Chapter 3
Self-reference and the divorce between meaning and truth

1 Introduction

Much contemporary work on the semantics of natural languages is guided by the assumption, inspired by Frege, that the meaning of a natural language sentence that is declarative in form is its truth condition. One kind of problem that that assumption (hereafter, "thesis F") has been claimed to face, and to which no generally accepted solution has so far been proposed, is that there exist declarative sentences that must have the same truth conditions even though they have evidently different meanings. Another kind of problem that it might face would be complementary to the above, and would consist in the existence of declarative sentences that have evidently identical meanings even though they must have different truth conditions. My purpose in this essay is to expose one problem of this second sort, on the basis of inter-linguistic and intra-linguistic evidence concerning the behaviour of a particular kind of self-referential sentence.

  A sentence is self-referential, as the term will be here understood, just in case it is the referent of one of its constituents. (Thus, the sentence "The sentence hereby formulated is interesting" is self-referential, since it is itself the referent of its constituent noun phrase "the sentence hereby formulated".) Liar-sentences (for example, "The sentence hereby formulated is false") are the most famous self-referential sentences in that sense, but none of those connected with the problem that I am about to describe are liar-sentences and, as I shall argue, the problem's significance derives in part from the fact that they are not.

2 What the problem is

If any statement of inter-linguistic synonymy is true, the following three certainly are:

   (a) The English noun phrase "the sentence hereby formulated" has the same meaning as the German noun phrase "der hiermit formulierte Satz".

    (b)   The English verb phrase "is an example of an English sentence" has the same meaning as the German verb phrase "ist ein Beispiel eines englischen Satzes".

    (c)   The English verb phrase "is an example of a German sentence" has the same meaning as the German verb phrase "ist ein Beispiel eines deutschen Satzes".

From these obviously true statements of inter-linguistic synonymy, together with standard assumptions about the compositionality of the processes by which meaningful noun phrases combine with meaningful verb phrases in forming meaningful sentences of English and German, it follows, first, that the sentences in (1) and (2) have the same meaning,

(1)   The sentence hereby formulated is an example of an English sentence.
(2)   Der hiermit formulierte Satz ist ein Beispiel eines englischen Satzes.

and, secondly, that the sentences in (3) and (4) have the same meaning:

(3)   The sentence hereby formulated is an example of a German sentence.
(4)   Der hiermit formulierte Satz ist ein Beispiel eines deutschen Satzes.

It is clear, however, that, although (1) and (2) do have the same *meaning*, they cannot have the same truth conditions: The referent of the noun phrase "the sentence hereby formulated" in (1) cannot be anything other than (1) itself, and (1) is consequently true just in case (1) is an example of an English sentence (which of course it is, hence (1) is true). On the other hand, the referent of the noun phrase "der hiermit formulierte Satz" in (2) cannot be anything other than (2) itself, and (2) is consequently true just in case (2) is an example of an English sentence (which of course it isn't, hence (2) is false). Similarly, although (3) and (4) have the same *meaning*, they cannot have the same truth conditions. For, the referent of the noun phrase "the sentence hereby formulated" in (3) cannot be anything other than (3) itself, and so (3) is true just in case (3) is an example of a German sentence (which of course it isn't, hence (3) is false); whereas the referent of the noun phrase "der hiermit formulierte Satz" in (4) cannot be anything other than (4) itself, and so (4) is true just in case (4) is an example of a German sentence (which of course it is, hence (4) is true). It is not the case, then, that whenever two declarative sentences have the same meaning, they have the same truth conditions. And if this is so, a declarative sentence's meaning cannot be identified with its truth condition, contrary to what thesis F contends.

It might be thought that this sort of problem would be difficult to arise at the intra-linguistic level. In fact, however, it is easy to ascertain its existence at that level, too.

For example, given that the noun phrase "the sentence hereby formulated" cannot be supposed to have a different *meaning* in each one of the different English sentences in which it might occur (though, of course, it will have a different *referent* in each one of the different English sentences in which it might occur), an adequate account of English that aims to compositionally derive the *meanings* of English sentences on the basis of the *meanings* of their constituents (and of the way these constituents are combined) would entail, among other things, that the sentences in (5) and (6) have the same meaning,

(5)   The sentence hereby formulated contains exactly eight words.
(6)   Exactly eight words are contained in the sentence hereby formulated.

that the sentences in (7) and (8) have the same meaning,

(7)   The sentence hereby formulated ends with a verb.
(8)   It is with a verb that the sentence hereby formulated ends.

and that the sentences in (9) and (10) have the same meaning:

(9)   The sentence hereby formulated isn't one that doesn't contain any contractions.
(10)   The sentence hereby formulated is not one that does not contain any contractions.

However, no adequate *truth* theory of English could generate such entailments. Although (5) and (6) do have the same meaning, they cannot have the same truth conditions: in (5), "the sentence hereby formulated" can only refer to (5) itself, and (5) is consequently true just in case it *itself* contains exactly eight words (which it does, hence (5) is true); in (6), on the other hand, "the sentence hereby formulated" can only refer to (6) itself, and (6) is consequently true just in case it *itself* contains exactly eight words (which it doesn't, hence (6) is false). Similarly, although (7) and (8) do have the same meaning, they cannot have the same truth conditions: what "the sentence hereby formulated" refers to in (7) can only be (7) itself, and so (7) is true just in case it *itself* ends with a verb (which it doesn't, hence (7) is false); on the other hand, what "the sentence hereby formulated" refers to in (8) can only be (8) itself, and so (8) is true just in case it *itself* ends with

a verb (which it does, hence (8) is true). Finally, although (9) and (10) do have the same meaning, they cannot have the same truth conditions. For, in (9), "the sentence hereby formulated" cannot refer to anything but (9), and so (9) is true just in case it *itself* contains at least one contraction (which it does, hence (9) is true); whereas, in (10), "the sentence hereby formulated" cannot refer to anything but (10), and so (10) is true just in case it *itself* contains at least one contraction (which it doesn't, hence (10) is false). In short, the inter-linguistic evidence falsifies just as clearly as the intra-linguistic evidence does the thesis that whenever two declarative sentences have the same meaning, they have the same truth conditions. And if that thesis is false, it cannot be true, as thesis F contends, that a declarative sentence's meaning is its truth condition.

3 What the problem is not

It might be claimed that the problem that sentences such as those cited above create for thesis F is but an instance of the well-known problem that sentences containing indexical expressions (hereafter, "indexical sentences") create for thesis F.

That claim does not withstand scrutiny, however. The problem that indexical sentences create for thesis F has two sources: first, that, considered in themselves (i.e. independently of their possible contexts of use), indexical sentences do not have *any* truth conditions; and secondly, that, considered in relation to their possible contexts of use, they can be associated with as many distinct truth conditions as there are relevantly distinct contexts in which they might be used (and so, in effect, with an infinite number of distinct truth conditions). For example, considered in themselves, the indexical sentences "Ich bin Deutsch" and "I am German" do not have any truth conditions; and when truth conditions are associated with them by reference to their possible utterers, each one of them gets associated with as many distinct truth conditions as there are distinct individuals who might utter it. As a result, if one were to identify, in accordance with thesis F, a sentence's meaning with its truth condition, one would be forced to absurdly conclude, with regard to indexical sentences such as "Ich bin Deutsch" or "I am German", either that they are meaningless or that each one of them has infinitely many distinct meanings. (Incidentally, the standard, Kaplanian, response to *that* problem involves abandoning thesis F. The meaning of an indexical sentence, in the Kaplanian framework, is its so-called "character", and an indexical sentence's character is definitely *not* its truth condition; see Kaplan [1989a, 1989b].)

Notice, however, that neither of the features that make indexical sentences troublesome for thesis F is a feature of the self-referential sentences responsible for the problem under discussion in this essay. For, first, none of these self-referential sentences lacks, considered in itself, a truth condition; and secondly, none of them is susceptible of being associated, when considered in relation to the infinite number of its possible contexts of use, with an infinite number of distinct truth conditions. Rather, each one of them has, considered in itself, a definite, unproblematically specifiable, truth condition; and that condition remains *the same* no matter what the context in which the sentence is used happens to be. So, the problem that thesis F has with these self-referential sentences *cannot* be that it wrongly entails, as it does wrongly entail for indexical ones, either that they are meaningless or that each one of them is associated with an infinite number of meanings. There is, consequently, no basis for supposing that the problem that the self-referential sentences considered above create for thesis F is an instance of the problem that indexical sentences are independently known to create for the same thesis.

One might grant that, for reasons such those just given, indexicality and self-referentiality should be acknowledged to be different things, but nevertheless develop the hope that, since the most famously problematic self-referential sentences are liar-sentences, the problem that the sentences considered in this essay create for thesis F will somehow turn out to be an instance of the problem that liar-sentences create for that thesis.

That hope would be groundless too, however. The problem that liar-sentences create for thesis F is that, although such sentences are meaningful, their meanings cannot be specified, as thesis F disposes one to suppose, by specifying their truth-conditions, since there can be no non-contradictory specification of their truth-conditions. And the source of that problem is not, of course, merely that liar-sentences refer to themselves, but, crucially, that what they *predicate* of themselves is a particular truth-theoretic property, namely the property of being false. (Indeed, it is arguable that self-referentiality not only is not sufficient but may not even be necessary for the generation of the kind of logically paradoxical result associated with liar-sentences; see Yablo [1993].)

Notice, however, that the self-referential sentences discussed in this essay are sharply different from liar-sentences in both of the above respects. First, each one of them has a truth condition whose specification is fully contradiction-free. And secondly, none of them predicates, of what it refers to, a *truth*-theoretic property of any kind (being a sentence of English, being a sentence of German, being a sentence of exactly eight words, being a sentence that ends with a verb, and being a sentence that contains at least one contraction, are certainly not *truth*-

theoretic properties of sentences in any intelligible sense of the term 'truth-theoretic'). It is therefore groundless to suppose that the problem that these sentences create for thesis F might turn out to be an instance of the problem that liar-sentences create for the same thesis.

4 Conclusion

Since there appears to be no chance of assimilating the problem examined in this essay to already known problems, I submit that it deserves its own distinctive place in the list of problems undermining thesis F.

To be sure, no problem is insurmountable if one is prepared to pretend that the evidence for its existence does not exist, and there are at least two ways in which one might attempt to do that in the present case. The simplest way would be to deny that (1)–(10) are sentences of English or German—in other words, to claim that a grammar of German would be descriptively *in*adequate if it generated (2) or (4), and that a grammar of English would be descriptively *in*adequate if it generated any of (1), (3) or (5)–(10). I take this claim to be too obviously wrong to merit any consideration.

A slightly more complex, though no less futile, way would consist in suggesting that there might be a sense of "meaning" in which the sentences that this essay assumes to be synonymous are not in fact synonymous. In response to such a suggestion, I would simply point out that, although the word "meaning" may have more than one senses, it has *none* in which the sentences in question could be held not to be synonymous. In particular, there is no sense of "meaning" in which the noun phrase "the sentence hereby formulated" acquires a different *meaning* (as opposed to a different referent) each time it appears in a different sentence, and there is no sense of "meaning" in which the noun phrase "the sentence hereby formulated" means anything different from what the noun phrase "der hiermit formulierte Satz" means; given these facts, and assuming that sentence meanings are determined compositionally on the basis of the meanings of sentential constituents, there is no way of avoiding the conclusion that the sentences that this essay assumes to be synonymous are indeed synonymous, despite the fact they *cannot* have the same truth conditions. Trying to ignore the trouble that this fact causes for thesis F by imagining senses of "meaning" that do not exist might be an understandable form of self-deception, but it would be a form of self-deception nonetheless.

Chapter 4
Performativity and the "true/false fetish"

> If you hurl a tomato at a political meeting (or bawl "I protest" when someone else does—if that is performing an action) the consequence will probably be to make others aware that you object…: but this will not make either the throw or the shout true or false (though they may be, even deliberately, misleading).
>
> Austin, *How to Do Things with Words*

1 Introduction

J. L. Austin took it to be a defining feature of the sorts of utterances that he came to call "explicit performative utterances" that they are neither true nor false, despite being utterances of declarative sentences, which are traditionally regarded as paradigms of truth-evaluability; and he did not feel compelled to give arguments for his non-truth-evaluability thesis, pointing out that he considers it to be too obvious to require defence. Wedded as they have been to a truth-conditional conception of linguistic content, subsequent philosophers of language have refused to accept Austin's non-truth-evaluability thesis—not surprisingly, since maintaining both it and the truth-conditional conception of content would force them to conclude that explicit performative utterances have no content at all. And since those philosophers, along with everyone else, do acknowledge both the contentfulness of explicit performatives and the significance of the phenomenon of performativity to which Austin was the first to pay systematic attention, they have sought to explain what is special about explicit performatives by devising accounts of them that not only do not incorporate Austin's non-truth-evaluability thesis but positively require precisely what Austin was ruling out—accounts, that is, according to which what is distinctive about explicit performatives cannot be understood *unless* they are taken to be bearers of a truth value.

The belief that such accounts are reliable, together with the absence of arguments on Austin's part, appears to be the main reason why Austin's thesis that explicit performatives are truth-valueless "is denied by almost everyone nowadays", as Hornsby (2006: 904) notes in an overview of post-Austinian work on performativity. But that denial would be apt for reconsideration, if it could be shown that anti-Austinian explanations of explicit performativity face insuperable problems that are due precisely to their assumption that explicit performa-

tives are truth-evaluable. This essay proposes to make a step towards such reconsideration. After a reminder of some characteristics of explicit performatives that are acknowledged by all parties to the dispute, I argue that attempts to justify the denial of Austin's non-truth evaluability thesis by producing explanations of performativity that essentially depend on the hypothesis that explicit performatives are truth-evaluable cannot succeed for at least two types of reason: on the one hand, because utterances that, on the proposed explanations, should be capable of being explicit performative ones turn out to be incapable of being explicit performative ones; on the other hand, because utterances that, on the proposed explanations, should be incapable of being explicit performative ones turn out to *be* capable of being explicit performative ones. Since the source of these explanatory failures turns out to be none other than the adoption of the hypothesis that explicit performatives are truth-evaluable, I suggest that they strongly undermine the anti-Austinian view and vindicate Austin's thesis, in favour of which I then sketch an independent argument based on the behaviour of explicit performatives in deductive inferential contexts (specifically, on the fact that their behaviour in such contexts could not be reconciled with the hypothesis that they are truth-evaluable unless one denied the applicability, in those contexts, of certain logically fundamental inference rules). My conclusion is that the Austinian thesis can by no means be regarded as having been superseded, and that Austin's opponents might even have to seriously consider adopting it if some of their own broader interests were to be safeguarded.

2 Performativity and the anti-Austinian view

The sorts of utterances to which Austin was focusing attention when he came to use the label "explicit performative utterances", and which are commonly discussed under that label today, are utterances of grammatically declarative sentences each of which is such that (a) its speaker, referring therein to himself/herself in the first person singular, predicates of himself/herself an illocutionary act named by a simple present tense, active main verb (and its object, if it has one), and (b) its issuance in the right circumstances constitutes the performance, by the speaker, of the act that he/she thereby predicates of himself/herself. Thus, an utterance, in the right circumstances, of the sentence, "I deny that arithmetic is complete", can constitute a speaker's denial that arithmetic is complete, and so can be an explicit performative utterance; whereas an utterance of the sentence, "I prove that arithmetic is incomplete", can under no circumstances constitute a

speaker's proof that arithmetic is incomplete, and so cannot be an explicit performative utterance. Similarly, an utterance, in the right circumstances, of the sentence, "I recommend the *Hammerklavier* sonata" can constitute a speaker's recommendation, to some hearer or hearers, and for some purpose or purposes, of the *Hammerklavier* sonata, and so can be an explicit performative utterance; whereas an utterance of the sentence, "I perform the *Hammerklavier* sonata", can under no circumstances constitute anyone's performance of the *Hammerklavier* sonata, and, therefore, cannot be an explicit performative utterance.

Two constraints on explicit performativity, both of them noted by Austin, are generally acknowledged and would be worth keeping in mind in the present context. The first is that a declarative sentence otherwise conforming to the type of declarative sentence described above is not, in general, performatively usable—i.e., usable in such a way that its production by its speaker can constitute the performance, by that speaker, of the act named by its active main verb—, if its active main verb occurs in a grammatical person other than the first person or in a grammatical tense other than the (usually, simple) present tense: Although "I request your support" can constitute my request of your support and "I offer you my car" can constitute my offer to you of my car (and so, can be explicit performative utterances), neither "I requested your support" nor "I offered you my car" can constitute my request of your support or my offer to you of my car, and so cannot be explicit performative utterances. Furthermore, although my saying to you "I request your support" can constitute my request of your support, and my saying to you "I offer you my car" can constitute my offer to you of my car, it is not the case that *your* saying to me "You request my support" can constitute *my* request of your support, or that *your* saying to me "You offer me your car" can constitute *my* offer to you of my car, which means that neither of these latter utterances can be an explicit performative utterance.

The second constraint on explicit performativity is that, even when a sentence fully conforms, as regards the grammatical features of its main verb, to the type of declarative sentence described above, it is, in general, only some, and not all, interpretations of those grammatical features that are compatible with the sentence's explicit performative use. In particular, if the present tense of the sentence's main verb is interpreted in a way that allows utterance-time and reference-time to diverge, the sentence cannot normally be used performatively. Thus, "I promise never to lie again" can constitute my promise never to lie again (and so, can be an explicit performative utterance), if the present tense of its main verb is understood as referring strictly to the time of speaking (as it would normally be, if the utterance was, for example, my response to your utterance of the imperative, "Promise me never to lie again!"). But it cannot constitute my promise

never to lie again, if, in uttering it, I am simply explaining to you what I do whenever I am caught lying (as in, "Each time I am caught lying, I promise never to lie again"); or if, in uttering it, I am simply describing to you the commitments I have undertaken in a letter that I have posted earlier today (as in, "In the final paragraph of the letter I have posted earlier today, I promise never to lie again"); or if, in uttering it, I am simply rehearsing what I plan to do in an upcoming important meeting ("Here is my plan for tonight's meeting: First, I thank them for accepting to see us. Second, I apologise for having lied yesterday. Third, I promise never to lie again.") It is, in short, only on specific interpretations of its grammatical features—and, in particular, on the (relatively uncommon) interpretation of the present tense as referring strictly to the time of speaking—that an utterance of a declarative sentence of the sort described above can normally be an explicit performative utterance.

Austin called "explicit performative verbs", or simply "performative verbs", the verbs that, under the constraints just noted, can be used as the main verbs of explicit performative utterances. And, in moving from the exposition of his doctrine of explicit performative utterances to the exposition of his doctrine of illocutionary acts, he suggested (Austin 1975: 149–150) that, if one wants to obtain a minimally comprehensive list of the types of illocutionary act that are commonly recognised in a language, there is no better method than to compile (as he himself had set out to do for that purpose) a list of the explicit performative verbs of that language—that is, a list of those verbs that can be used as the main verbs of explicit performative utterances whose issuance in appropriate circumstances would constitute the acts that the verbs name.[1]

---

1 In expounding his doctrine of illocutionary acts, Austin occasionally used the term "performative" not merely as a shorthand for "explicit performative", but also in a different, extended, sense, in which it refers to *any* utterance accomplishing some speech act or other, whether or not the utterance *names* (in the way his "explicit performative utterances" do) the act it accomplishes. Coupled with Austin's thesis that every natural language utterance normally accomplishes some speech act or other, this terminological choice has the unfortunate consequence that every normal natural language utterance is "performative" in the extended sense, and thus deprives the notion of performativity in the extended sense of any clearly distinctive theoretical role. In this essay I use "performative" only as a shorthand for Austin's "explicit performative"—that is, only by reference to utterances that name (in their first-person present-tense active main verbs) the illocutionary acts they accomplish—, and not in Austin's inflationary extended sense; similarly for "performativity".

It seems that Austin became aware of the problem created by his occasional use of "performative" in the extended sense when it was too late for him to correct it: the editors of *How to Do Things with Words* tell us in their Appendix that, at the point in Austin's lecture notes where the transition to the doctrine of illocutionary acts (and to the inflationary use of "performative") is

Apart from their use in providing him with what he regarded as an appropriate point of entry into his nascent theory of illocutionary acts, the main use that Austin has made of explicit performative utterances (and the one that is our primary concern here) was a polemical one. For, Austin took it to be a defining feature of explicit performative utterances not only (i) that their speakers, by issuing them in appropriate circumstances, accomplish the illocutionary acts that they thereby predicate of themselves, but also (ii) that, in predicating those acts of themselves, their speakers do *not* produce, nor do they intend to produce, a truth-evaluable representation of themselves as accomplishing the acts in question (and so, cannot be supposed to be accomplishing them by way of describing themselves as accomplishing them, or by way of stating that they are accomplishing them, or by way of declaring that they are accomplishing them, or by any other way that would require of them to be producing, in accomplishing them, a truth-evaluable representation of themselves as accomplishing them). However, explicit performative utterances are utterances of grammatically declarative sentences; and since, on the view that was prevalent in the philosophy of Austin's time, utterances of grammatically declarative sentences (if such sentences did not belong to those that were condemned as nonsensical for positivistic reasons) were assumed to be, and to be intended to be, truth-evaluable representations of reality, the existence of explicit performative utterances constituted, for Austin, a distinctive and decisive type of evidence against the aforementioned assumption, which he dubbed "the descriptive fallacy" and against which he has found additional occasions to position himself in his work.[2]

---

effected, there is "a marginal note dated 1958" in which Austin writes, "All this isn't clear!" and asks rhetorically, "Won't all utterances be performative?" (Austin 1975: 167). Unlike Austin, some of his commentators, and all of his popularisers, appear never to suspect that the use of "performative" in the extended sense risks trivialising the notion of performativity; what is worse, they sometimes advance arguments where conclusions about performativity in the strict sense (that is, explicit performativity) are fallaciously drawn from premises concerning performativity in the extended sense, or conversely.

2 Austin's non-truth evaluability thesis regarding explicit performatives is asserted in all three of his extended treatments of them that have followed his brief, incidental, discussion of the topic in his 1946 article "Other Minds": the 1955 Harvard lectures posthumously published as *How to Do Things with Words* (Austin 1975), the 1956 BBC talk posthumously published as "Performative Utterances" (Austin 1979: 233–252), and the 1958 Royaumont Abbey talk posthumously published as "Performatif-Constatif" and translated as "Performative-Constative" (Austin 1963). On all three of these occasions, the non-truth-evaluability of explicit performatives is presented as evidence against what Austin calls the "descriptive fallacy", a view that he had already targeted under that name in "Other Minds" (1946: 174; 1979: 103). The non-truth evalu-

Post-Austinian philosophers of language are, if anything, even more unwilling than some of Austin's own contemporaries might have been to accept Austin's claim that explicit performative utterances are truth-valueless, since most of them have come to subscribe to truth-conditional theories of linguistic content; if Austin were right, therefore, their accounts of linguistic content should be acknowledged to have the undesirable consequence that they constrain them to claim, about certain obviously contentful utterances of natural language sentences, that they have no content at all. On the other hand, post-Austinian philosophers of language would have even less an excuse than Austin's contemporaries might have to claim that the truth-evaluability, and hence (for them) the contentfulness, of explicit performatives simply *follows* from their declarative grammatical form, since that claim, besides begging the question against Austin, would beg the question against grammatical theory itself: whatever else a grammar of a natural language is, it is nowadays widely agreed that it is *not* something that is capable of delivering verdicts as to which linguistic objects are truth-evaluable and which aren't. It is therefore not surprising that those post-Austinian philosophers of language who have attempted to seriously address the issue raised by Austin should have sought to resist his non-truth-evaluability claim about explicit performatives in a way that, at least in appearance, neither begs the question against Austin nor makes gratuitous assumptions about imaginary pronouncements of grammatical theory. And the way that most of them have found to be adequate to that task can best be represented as a kind of abductive argument in favour of the denial of Austin's claim—an argument, specifically, that purports to justify the *hypothesis* that explicit performatives are truth-evaluable by claiming that, if one assumes, *contra* Austin, that they are truth-evaluable, then one can *explain* the most distinctive feature of explicit performatives that

---

ability of explicit performatives is also asserted, and presented as evidence against the "descriptive fallacy", in Austin's 1950 article "Truth", which mentions other kinds of utterances of declarative sentences besides explicit performatives that, in Austin's view, are not truth-evaluable despite their declarative grammatical form (1950: 125–127; 1979: 130–132).

It may be worth noting that Austin's teacher, H. A. Prichard, in a paper on promising written, according to its first editor, circa 1940, and published posthumously in 1949 (see now Prichard 2002: 257–265), explicitly associates what Austin was later to call explicit performativity with non-truth-evaluability. Commenting on the promise made in uttering "I promise not to reduce the rates", Prichard writes that "while everyone would allow that a promise may be made either in good or in bad faith, no one would allow that it could be either true or false", and adds that "promising resembles asking a question or issuing an order in that it consists not in making a statement" (Prichard 2002: 258).

both Austin and everyone else acknowledges, namely that their speakers can accomplish, in uttering them, the acts that they thereby name. According to the proposed explanation, an explicit performative utterance can accomplish, by being issued, the act that it names because it is a truth-evaluable utterance of a declarative sentence which, unlike truth-evaluable utterances of other types of declarative sentences, has the special property that its truth-condition is such that it can be satisfied, thus rendering the utterance true, by the utterance's own issuance. (Thus, what accounts for the fact that an utterance like "I deny that arithmetic is complete" can be an explicit performative utterance—that is, can constitute a denial that arithmetic is complete—whereas an utterance like "I prove that arithmetic is incomplete" cannot be an explicit performative utterance—that is, cannot constitute a proof that arithmetic is incomplete—is that, although they both are truth-evaluable utterances of declarative sentences, the truth condition of the former, unlike the truth condition of the latter, is such that it can be satisfied, thus making the utterance true, by the utterance's own issuance). And, according to the abductive argument that I am here reconstructing, it is precisely because it affords this explanation of what explicit performativity really is that the hypothesis that explicit performatives are truth-evaluable is justified, and can be upheld in opposition to Austin's thesis that they aren't.

In a section of his *Philosophical Analysis in the Twentieth Century* entitled "A Lesson about Explicitly Performative Sentences", Scott Soames expresses this currently widespread anti-Austinian understanding of explicit performativity in these terms:

> [T]he explicitly performative sentence *I promise to return the book* has a straightforwardly descriptive meaning that represents the world as being a certain way, and so imposes truth conditions on it. What makes this sentence special, and performative, is that one can bring it about that these truth conditions are satisfied (i.e., one can bring it about that the proposition expressed by the sentence is true) simply by uttering the sentence in the right circumstances. (Soames 2003: 127)

Soames mentions Lewis (1979) as an early exponent of a version of this view, but an even more influential twentieth century philosopher, Quine, could have been mentioned as well. In his contribution to a symposium on Austin held in 1965 and first published in its entirety in 1969, Quine acknowledges that Austin's discussion of explicit performatives utterances in *How to Do Things with Words* was directed against what Austin there calls—identifying it as one of the principal targets of his work—the "true/false fetish" (Austin 1975: 151), but claims that explicit

performative utterances do not justify "Austin's animus against the true/false fetish", since what is special about an explicit performative utterance—in Quine's view—is not, as Austin thought, that it lacks a truth condition, but rather that its truth condition is such that it can be satisfied by the utterance's own issuance, thus rendering the utterance one that is true:

> "I bid you good morning" is true of us on a given occasion if and only if, on that occasion, I bid you good morning. A performative is a notable sort of statement, I grant; it makes itself true; but then it is true. (Quine in Urmson, Quine and Hampshire 1969: 90)

Variously embellished, the view that Quine here expresses underlies almost all post-Austinian accounts of explicit performatives, including the two currently most popular (and otherwise antagonistic) ones, due respectively to Kent Bach (1975) and to John Searle (1989). According to Bach, explicit performative utterances are "statements" that "are true in virtue of being made" (Bach 1975: 230); and according to Searle, "performative utterances are those in which saying something makes it true" (Searle 1998b: 115).

I will call the kind of view just outlined "the anti-Austinian view", and it will be the object of critical attention in what follows. It should be noted that although all proponents of the anti-Austinian view agree that explicit performative utterances are truth-evaluable utterances of declarative sentences whose truth conditions are such that they can be satisfied, thus making the utterances true, by the utterances' own issuance, they need not agree, and in fact they rarely do agree, on the question as to *why* explicit performatives utterances happen to have (whereas truth-evaluable utterances of other types of declarative sentences happen not to have) truth conditions of that special kind. Since examining, on a case by case basis, the conflicting answers to that question given by various anti-Austinians would be relevant only if one had no doubt about the correctness of the basic anti-Austinian claim to which they all subscribe, and since it is the correctness of that basic claim that I doubt, I will have nothing further to say about internal disputes within the anti-Austinian camp in this essay.

3 Against the anti-Austinian view, Part I

Suppose, as the anti-Austinian view has it, that the explicit performativity of an utterance (that is, the fact that it can constitute the act that it names) is due to its

having a truth condition that is such that it can be satisfied, thus making the utterance true, by the utterance's issuance. It follows from this that any utterance that has *the same truth condition* as the truth condition that, by hypothesis, a given explicit performative utterance has will also be capable of constituting, through *its* issuance, the act that it names: if the capacity of an utterance to be used performatively is correctly explicable as resulting from its truth-condition's being such as to be satisfiable by the utterance's issuance, then issuances of any utterances that *share* an explicit performative's supposed truth condition should also share its capacity to be used performatively.

Call this consequence of the anti-Austinian view Thesis T. Among the ways of seeing that Thesis T is false, and that, therefore, the anti-Austinian view cannot be right, a simple one consists in assuming the nowadays standard, Kaplanian, view of the interpretation of indexicals, and in considering the effect of replacing, in certain contexts, the first-personal indexical subject term of an explicit performative utterance with a co-referential second-personal indexical subject term. The relevant part of the standard view of the interpretation of indexicals is simply the thesis that singular indexicals are directly referential expressions, and can be expressed as follows: given a declarative sentence containing a singular indexical, the indexical's contribution to the truth condition that, relative to a context of use, the declarative sentence has is *just* the referent of the indexical in that context of use, rather than any 'way of presenting' or any 'way of fixing' that referent in that context of use (notice that the standard view does not deny that different 'ways of presenting' or different 'ways of fixing' their referents may be associated with distinct singular indexicals as parts of their distinct linguistic meanings; what it asserts is that it is *only* their referents, and not those 'ways of presenting' or 'ways of fixing' their referents, that the indexicals contribute to the context-relative *truth conditions* of the sentences where they occur; cf. Kaplan 1989a, Kaplan 1989b). It follows from this that if the *only* difference between two truth-evaluable declarative sentences is that they contain distinct singular indexicals that are co-referential relative to a context of use, then the two sentences have exactly *the same truth condition* relative to that context of use. For example, relative to a context in which John produces a token of (1) while Maria, addressing John, produces a token of (2),

(1)   I am Italian.
(2)   You are Italian.

(1) and (2) have exactly the *same* truth condition, namely the condition that John is Italian; and relative to a context in which Maria produces a token of (1) while

John, addressing Maria, produces a token of (2), (1) and (2) have exactly the same truth condition, namely the condition <u>that Maria is Italian</u>.

But now, if the standard view of the interpretation of indexicals is taken to apply to explicit performative utterances, as it must be taken to apply *if* explicit performatives are assumed to be truth-evaluable, then the anti-Austinian view is constrained to derive the obviously false conclusion that a host of utterances that are clearly incapable of being explicit performatives ones are in fact capable of being explicit performatives ones, since those utterances will have exactly the same truth conditions as the explicit performative ones, and so their issuance will be capable of satisfying those conditions just as much as the issuance of the explicit performative ones does. Consider a context in which John produces a token of (3) while Maria, addressing John, produces a token of (4), and in which the present tense markers of both John's and Maria's tokens have the same present time reference:

(3)   I deny that arithmetic is complete.
(4)   You deny that arithmetic is complete.

Assume—as it can certainly be the case—that John's token is an explicit performative one (that is, in producing it, John accomplishes the act, which he names, of denying that arithmetic is complete). Then, given the standard view of the interpretation of indexicals, the anti-Austinian view entails the obvious falsehood that not only John's but also *Maria's* token can constitute *John's* denial that arithmetic is complete. For, on the assumption that John's explicit performative token is truth-evaluable, the standard view of the interpretation of indexicals will entail that John's and Maria's tokens have exactly *the same* truth condition—namely, the condition <u>that John denies that arithmetic is complete</u>. And on the further assumption that a token's explicit performativity is due to its having a truth condition that is such that it can be satisfied by the token's issuance, it will follow that the issuance of *either one* of the two tokens can satisfy their identical truth condition, and can thus be an explicit performative token—it will follow, in other words, not only that John's token of (3) can constitute his denial that arithmetic is complete, but also that Maria's token of (4) can constitute *John's* denial that arithmetic is complete. Since, however, it is obvious that, although John's token of (3) can constitute John's denial that arithmetic is complete, Maria's token of (4) cannot constitute John's (or, for that matter, anyone else's) denial that arithmetic is complete, the conclusion must be that the anti-Austinian view cannot explain the explicit performativity of (3) without falsely attributing explicit performativity to (4).

This point is obviously generalisable to indefinitely many cases, indicating that the explanatory failure of the anti-Austinian view that it signals is a failure on a massive scale. Take *any* pair of utterances of present tense declarative sentences of the forms "I *Φ*" and "You *Φ*", naming an illocutionary act of *Φ*-ing, of which the first can be an explicit performative utterance (that is, can *constitute* the act of *Φ*-ing that it names) while the second cannot be an explicit performative utterance (that is, cannot constitute the act of *Φ*-ing that it names), and in which the present tense forms have the same present time reference and the first- and second-person singular indexicals are co-referential. Then, on the anti-Austinian assumption that the utterance of the form "I *Φ*" is truth-evaluable, the standard view of the interpretation of indexicals will entail that "I *Φ*" and "You *Φ*" have exactly the same truth condition. And this, together with the further anti-Austinian assumption that the performativity of an utterance is due to its having a truth condition that is such that it can be satisfied by the utterance's issuance, will entail the obvious falsehood that "You *Φ*" can be an explicit performative utterance just as much as "I *Φ*" can—in other words, will entail that the issuance of "You *Φ*" can constitute the act of *Φ*-ing that it names just as much as the issuance of "I *Φ*" can constitute the act of *Φ*-ing that it names.

In order to avoid this explanatory collapse while remaining true to his or her anti-Austinianism, the anti-Austinian would have to defend the extraordinary claim that an explicit performative utterance of the form "I *Φ*", even though it is, by anti-Austinian lights, truth-evaluable, *cannot* have the same truth condition as a corresponding non-performative utterance of the form "You *Φ*", when the present tense forms of both utterances have the same present time reference, and their first- and second-person indexicals are co-referential. But it is very hard to see how this claim could be credibly defended, since its defence would require revisionary assumptions, which hardly anyone would be willing to accept, about either the interpretation of indexical expressions or the interpretation of performative verbs. If the claim's defence were meant to rely on a revisionary assumption about the interpretation of indexical expressions (rather than of performative verbs), the claim would be rejected, since the requisite assumption would have to be nothing less than the assumption that it is impossible for the members of a set of truth-evaluable declarative sentences to have the same truth condition, relative to a context of use, when those members differ *only* in that they contain distinct singular indexicals that are co-referential relative to that context of use. But taking this to be impossible entails denying that, for example, "I am fifty years old", said by me to you today, can have the same truth condition as "You are fifty years old," said by you to me today. And any theory of indexicality that is constrained to deny *that* would most likely be rejected.

If, on the other hand, the defence of the anti-Austinian claim were to rely on a revisionary assumption about the interpretation of performative verbs (rather than of indexical expressions), then it would be rejected too, since the requisite assumption about the interpretation of performative verbs would have to be that every apparently unambiguous performative verb is ambiguous between at least two necessarily divergent senses, in the first of which the verb has *only* first-person present-tense occurrences and in the second of which it has *only* non-first-person non-present-tense occurrences (on that account, that is, the reason why, for example, "I deny that arithmetic is complete", said by me to you, and "You deny that arithmetic is complete", said by you to me, would, allegedly, have necessarily disjoint truth conditions would be that the verb 'deny', as used in the first utterance, is *not* the same verb as, and has necessarily a different meaning from, the verb 'deny' as used in the second utterance). But the assumption that every apparently unambiguous performative verb is ambiguous in this highly peculiar way, besides being one that every anti-Austinian is on record as emphatically denying, is one that would entail that explicit performative utterances, as normally understood, do not exist at all, and would, if only for that reason, be rejected. For, on the normal understanding of performativity, an explicit performative utterance of the form "I Φ" is, among other things, one whose issuance by a speaker would have satisfied an imperative utterance of the form "Φ!" that would have been addressed to that speaker (if you were to address to me the imperative "Promise never to lie again!", my utterance of the performative "I promise never to lie again" would have satisfied your imperative; if you were to address to me the imperative "Admit that you are guilty!", my utterance of the performative "I admit that I am guilty" would have satisfied your imperative; and so on). On the assumption under consideration, however, no explicit performative of the form "I Φ" could ever satisfy a corresponding imperative of the form "Φ!", since, despite all appearances, the act that the utterer of the performative would be naming and performing would, of necessity, *not* be the act whose performance the utterer of the imperative would have solicited. And this is surely a consequence that no account of performatives (or, for that matter, of imperatives) would tolerate.

Since the anti-Austinian has no credible way of avoiding the explanatory collapse engendered by the falsity of Thesis T, the claim from which Thesis T follows—namely, that the explicit performativity of an utterance is due to its having a truth condition that is such that it can be satisfied by the utterance's issuance—must be rejected. And there are, of course, many other types of case that lead to the same kind of explanatory collapse. One type of case that is particularly strik-

ing (and indicates that trying to make the anti-Austinian explanation more restrictive would not remove its explanatory impotence) becomes apparent if one considers the effect of replacing, in certain kinds of context, the indexical subject term of an explicit performative utterance with a co-referential proper name, and if one assumes (as it is very widely assumed) that, at least in that sort of syntactic position, proper names are as much directly referential expressions as indexicals are (that is, contribute *only* their referents to the truth conditions of the sentences containing them). Suppose that Steven Stevens is amnesiac about his name, and that, fully believing that he is someone *other* than Steven Stevens, he makes what he takes to be a conjecture about Stevens' current deeds by saying,

(5)   Steven Stevens denies/bets/concludes that there is life in other planets.

It is clear that Stevens would *not* have thereby denied/bet/concluded (unbeknownst to himself, as it were) that there is life in other planets. On the other hand, if the amnesiac Stevens were to say instead,

(6)   I deny/bet/conclude that there is life in other planets.

he *would* have thereby denied/bet/concluded that there is life in other planets. The anti-Austinian view, however, cannot acknowledge the explicit performativity of (6) without falsely attributing explicit performativity to (5). For since, if both (5) and (6) were truth-evaluable, they would, in the circumstances, have exactly the same truth condition, the anti-Austinian view would wrongly entail that the issuance of (5) could constitute the act of denying/betting/concluding that it names just as much as the issuance of (6) could.

Indeed, it is not even necessary to invoke atypical conditions like a *speaker's* ignorance of his or her own name in order to produce cases with just this effect. Far from atypical conditions like an *addressee's* ignorance of a speaker's name would do as well. Suppose that I am Philip Philips and know full well what my name is, but that *you* don't know what my name is, and I am fully aware of the fact that you don't know what my name is. If, in these circumstances, I were to say to you,

(7)   I condemn/applaud your actions.

I *would* have thereby condemned/applauded your actions. But if, in the same circumstances, I were to say to you,

(8)   Philip Philips condemns/applauds your actions.

I would *not* have thereby condemned/applauded your actions, no matter what I might be 'privately' taking myself to be doing. However, if both (7) and (8) were truth-evaluable, then, relative to a context in which they would be uttered by me and addressed to you, they would have exactly the same truth condition; and this, given the anti-Austinian explanation of performativity, would lead to the obviously false conclusion that my utterance of (8) could constitute my act of condemning/applauding your actions just as much as my utterance of (7) could.

It is important to bear in mind, in considering all these counterexamples, that the anti-Austinian explanation of performativity is best viewed as purporting to enable an *abductive* argument in favour of the anti-Austinian thesis that performatives are truth-evaluable: It is not as if the anti-Austinian thesis were somehow *obviously* true, and should be accepted *whether or not* it led to an adequate explanation of the phenomenon of performativity; rather, its leading to an adequate explanation of the phenomenon of performativity would provide the principal reason for thinking that it *is* true; consequently, its *not* leading to such an adequate explanation removes the principal reason for taking it to *be* true, and thus for undermining Austin's thesis. Indeed, viewed from such an abductive perspective, Austin's contrary thesis that explicit performatives are truth-valueless would now appear to offer the simplest explanation of the fact, revealed in the above discussion of the counterexamples, that certain non-first-personal utterances that should behave exactly as explicit performatives utterances do, *if* the latter were truth-evaluable in the way the anti-Austinian view supposes, do not so behave: they do not so behave, the Austinian would hold, because they *are* truth-evaluable utterances, whereas explicit performatives *aren't*.

4 Against the anti-Austinian view, Part II

The argument of the previous section was that anti-Austinian explanations of performativity cannot be correct since there exist utterances that, on those explanations, should be capable of being used performatively, but nevertheless turn out to be incapable of being used performatively. The argument of the present section will be that anti-Austinian explanations cannot be correct for the complementary reason that there exist utterances that, on those explanations, should be incapable of being used performatively, but nevertheless turn out to *be* capable of being used performatively.

Suppose again, as the anti-Austinian view proposes, that the explicit performativity of an utterance (that is, the fact that it can constitute the act that it names) is due to the fact that its truth condition is such that it can be satisfied, thus making the utterance true, by the utterance's own issuance. From this it follows that nothing could be an explicit performative utterance if, considered as something truth-evaluable, it would entail something that happens to be false. For, if something that is truth-evaluable entails something that happens to be false, it cannot itself be true. And if something cannot be true, it cannot be true by virtue of being issued. So, if an explicit performative utterance is something truth-evaluable that is true by virtue of being issued, then nothing can be an explicit performative utterance if, considered as something truth-evaluable, it entails something that happens to be false.

Let us call this consequence of the anti-Austinian view Thesis F. There are several types of counterexample showing that Thesis F is false, and that, therefore, the anti-Austinian view cannot be correct, but I shall here concentrate on a particular type of case, which is connected to the important topic of the role of the concepts of sincerity and insincerity in accounts of linguistic action. Consider the following utterances,

(9)   I sincerely assert that I haven't been here before.
(10)   I sincerely wish you a speedy recovery.
(11)   I sincerely apologise for my last remark.
(12)   I sincerely promise to pay all my debts to you.

concerning which the following statements indisputably hold. First, each one of them can constitute, by virtue of being issued, the act (of asserting, of giving one's wish, of apologising, and of promising, respectively) named by its first-person present-tense main verb—in other words, each one of them can be an explicit performative utterance. Second, each one of them can constitute the act named by its first-person present-tense main verb (and so, can be an explicit performative utterance) *even when its utterer does not satisfy the characteristic sincerity expectation* associated with performances of the act in question—that is, even when its utterer does not believe what he asserts (in the case where the act he names and performs is an assertion), does not desire or hope what he is giving his wish for (in the case where the act he names and performs is the act of giving his wish), does not feel regret over what he apologises about (in the case where the act he names and performs is an apology), and does not intend to do what he promises to do (in the case where the act he names and performs is a promise). Thus, I can successfully assert, in uttering (9), that I haven't been at the place of my utterance

at a time prior to the time of my utterance, even though I do not believe that this was the case, and indeed even if I *know* that it was not the case (I may simply, in making my assertion, be lying to a police officer who is conducting a murder investigation); similarly, I can successfully, in uttering (10), give you my wish for your speedy recovery, even though I don't really want you to recover soon, and indeed even if I secretly hope that you will *not* recover soon; again, I can apologise, in uttering (11), for a previous remark of mine, even though, far from feeling the slightest regret about having made it, I am secretly delighted about having made it, noticing that it has had the damaging effect that I had intended; and of course, I can promise, in uttering (12), to pay all my debts to you, even though I do not have the intention of paying any of my debts to you, but am simply trying to gain time in order to better organise my disappearance.

The phenomenon that these cases exemplify is perfectly general, and it is one that no adequate account of linguistic action should fail to recognise: for any illocutionary act that generates a characteristic sincerity expectation (in the case of assertions, the expectation that the speaker believes what she asserts; in the case of apologies, the expectation that the speaker regrets what she apologises about; in the case of promises, the expectation that the speaker intends to realise what she promises; and so on), a speaker can successfully perform the illocutionary act—by, among others, explicit performative means, when such are available—*without* satisfying its associated sincerity expectation. And, of course, serious attempts at constructing theories of illocutionary acts routinely recognise, under various terminologies, this fact. For example, Searle and Vanderveken, who call the characteristic sincerity expectation generated by an illocutionary act the "sincerity condition" on that act, rely, in their *Foundations of Illocutionary Logic*, on a fundamental distinction between, on the one hand, "success conditions" on illocutionary acts (that is, conditions whose non-satisfaction entails the *non*-performance of the acts) and, on the other hand, "non-defectiveness conditions" on illocutionary acts (that is, conditions whose non-satisfaction does *not* entail the non-performance of the acts), and clearly acknowledge that all so-called "sincerity conditions" on illocutionary acts are merely conditions on illocutionary non-defectiveness and *not* conditions on illocutionary success (Searle and Vanderveken 1985: Chapter 1). And although Searle and Vanderveken misleadingly suggest (Searle and Vanderveken 1985: 13) that Austin "fails" to draw such a distinction, the truth is that Austin makes an exactly parallel distinction, and an exactly parallel acknowledgment, in his long discussion, in *How to Do Things with Words* (Austin 1975: Lectures II-IV), of the conditions under which an explicit performative utterance naming an illocutionary act may be, in his terms, "infelicitous" with respect to the performance of the act that it names. For, Austin

there sharply distinguishes between two fundamental types of such conditions (and, correspondingly, two fundamental types of "infelicity"): on the one hand, those whose non-satisfaction (called by him a "misfire") entails the *non*-performance of the act named by the performative verb, and, on the other hand, those whose non-satisfaction (called by him an "abuse") does *not* entail the non-performance of the act named by the performative verb; and he explicitly assigns all conditions having to do with characteristic sincerity expectations associated with performatively usable illocutionary verbs (these are the conditions belonging to sub-class "Γ.1." in his classification of felicity conditions) to the category of conditions whose non-satisfaction would merely constitute an illocutionary "abuse" and *not* an illocutionary "misfire".

But although it is evidently true, and widely recognised, that an illocutionary act denoted by a performative verb can be accomplished even when its agent does not satisfy the characteristic sincerity expectation associated with that act, the implications that this phenomenon has on the analysis of utterances such (9)–(12) have not been investigated, and are especially significant in the present context. For, not only can speakers of these utterances accomplish the acts they name without satisfying their associated sincerity expectations, they can accomplish them in that way *in spite of the fact* that they present themselves as *satisfying* those expectations. And this means that, *considered as truth-evaluable utterances*, these utterances could not fail to be *false* when they thus accomplish the acts they name and accomplish.

The key fact to notice in order to appreciate this point is that, although the bare statement that an agent performs a certain illocutionary act does not entail (as just noted) that the agent satisfies the sincerity expectation associated with that act, the statement than an agent *sincerely* performs a certain illocutionary act *does* entail that the agent satisfies the sincerity expectation associated with that act—that is why, for example, third-person statements like (13)–(16) are not contradictory, whereas corresponding third-person statements like (17)–(20) *are* contradictory:

(13)   He asserted that he hadn't been there before, even though he didn't believe that this was the case, and indeed knew that it wasn't the case.

(14)   He wished her a speedy recovery without really wanting her to recover.

(15)   He apologised for his remarks, even though he was not feeling any regret about those remarks.

(16)   He promised to pay all his debts to her, without having the slightest intention of paying any of his debts to her.

(17)   He sincerely asserted that he hadn't been there before, even though he didn't believe that this was the case, and indeed knew that it wasn't the case.

(18)   He sincerely wished her a speedy recovery without really wanting her to recover.

(19)   He sincerely apologised for his remarks, even though he was not feeling any regret about those remarks.

(20)   He sincerely promised to pay all his debts to her, without having the slightest intention of paying any of his debts to her.

Once this point is appreciated, however, it is not difficult to see that the undeniable fact that speakers of explicit performative utterances such as (9)–(12) can be performing the illocutionary acts they name without satisfying the sincerity expectations that they therein *present themselves* as satisfying is simply inconsistent with the anti-Austinian explanation of performativity. Suppose that, anxious to avoid the possibility of being enlisted as a murder suspect, I produce, responding to a relevant inquiry by a police officer, the following utterance:

(9)   I sincerely assert that I haven't been here before.

My utterance of (9) would certainly constitute the *assertion* that it names—namely, the assertion that I haven't been at the place of my utterance at a time prior to the time of my utterance—, even though I might not, while uttering it, *believe* that I haven't been at the place of my utterance at a time prior to the time of my utterance, and might even know that I *have* frequently been there in the past. (Notice that if the police officer later discovers that I have frequently been there in the past, he could rightly accuse me of having *asserted*, falsely, that I have not been there in the past; and that I could hardly defend myself against that accusation by claiming that, since I never *believed* that I hadn't been there in the past, I have *not* made that assertion at all, and so cannot have made it falsely.) On the anti-Austinian account, however, if I were to utter (9) without believing that I haven't been at the place of my utterance at a time prior to the time of my utterance, I *could not* thereby be *asserting* that I haven't been at the place of my utterance at a time prior to the time of my utterance, and my utterance of (9) *could not* have been an explicit performative utterance. For, on the anti-Austinian view, an explicit performative utterance is a truth-evaluable utterance that is rendered *true* by being issued. Considered as a truth-evaluable utterance, however, my utterance of (9) would, in the envisaged circumstances, be certainly *false*, since what it would *entail*—namely, that *I believe* that I haven't been at the

place of my utterance at a time prior to the time of my utterance—would itself be false. And since an utterance that is false cannot at the same time be true, let alone true in virtue of being issued, my utterance of (9) would wrongly be taken, on the anti-Austinian view, to be an utterance that could not possibly be an explicit performative utterance.

For exactly analogous reasons, the anti-Austinian view makes obviously incorrect predictions about a host of relevantly similar cases. For example, in uttering (10), (11) and (12),

(10)   I sincerely wish you a speedy recovery.
(11)   I sincerely apologise for my last remark.
(12)   I sincerely promise to pay all my debts to you.

I may, respectively, be *giving you my wish* for your speedy recovery without really *wanting* you to recover, be *apologising* for a previous remark of mine without really *regretting* that remark, and be *promising* to pay my debts to you without in the least *intending* to pay my debts to you. On the anti-Austinian view, however, none of these things could possibly happen; for, (10), (11) and (12) cannot, on that view, be used for performing the illocutionary acts they name unless they are truth-evaluable utterances that are rendered *true* by being issued; and since, considered as truth-evaluable utterances, (10), (11) and (12) would, in the envisaged circumstances, be *false* rather than true—given that what they would respectively *entail* about my desires, my regrets, and my intentions would be false—it would follow, contrary to fact, that (10), (11) and (12) could not possibly be, respectively, my wish for your speedy recovery, my apology for my remark to you, and my promise to pay my debts to you. Indeed, there is literally no end to the wrong predictions that the anti-Austinian view would commit its adherent to in this area: For *any* illocutionary act that is associated with a characteristic sincerity expectation, and that a speaker performs through an explicit performative utterance in which he or she *purports* to be satisfying the act's associated sincerity expectation without in fact satisfying it, the anti-Austinian view will wrongly entail that the illocutionary act itself is not performed at all, and hence that what is in fact an explicit performative utterance is not an explicit performative utterance.

There is exactly one way in which the anti-Austinian might try to remain anti-Austinian in view of this new explanatory impasse, but I don't believe that anyone, including the anti-Austinian, would, on reflection, be willing to employ it: It would consist in stipulating that nothing is to count as an illocutionary act unless its associated sincerity expectation is in fact satisfied (thus, no one is to count

as asserting something unless one believes what one asserts, no one is to count as apologising for something unless one feels regret over what one apologises about, no one is to count as promising something unless one intends to do what one promises to do, and so on). The problem with this stipulation, of course, is that it would oblige the anti-Austinian to define out of existence a great many things that without any doubt exist—for example, lies. Suppose that you are suspected of having committed a crime, and that, in response to a relevant inquiry, you defend yourself by saying:

(21)   I sincerely assert that I am innocent.

Everyone would agree that, in saying this, (a) you would be *asserting* that you are innocent, and (b) you might be *lying* about your innocence. According to the stipulation under consideration, however, that would be quite impossible: if your utterance *is* an assertion that you are innocent, then it *cannot* be a lie about your innocence, since assertions that can be lies simply *do not exist*. This is the sort of consequence that, I suppose, would force everyone, including the anti-Austinian, to reject the stipulation under consideration as a defensible way of saving the anti-Austinian view from its new explanatory impasse. And since no other decent way appears to be available, no salvation appears to be forthcoming.

Since what causes this explanatory impasse is the falsity of Thesis F, the claim from which that thesis follows—namely, that an explicit performative utterance is a truth-evaluable utterance that is rendered true by virtue of being issued—must be rejected. And if that claim is rejected, it can of course provide no motivation whatsoever for the anti-Austinian view that explicit performative are truth-evaluable. On the contrary, it is the Austinian view that performatives are *not* truth-evaluable that, if accepted, would help one understand why there has been an impasse here in the first place: if explicit performatives are not truth-evaluable, then they do not literally *have* entailments, and so the question of what happens to 'their' truth values when 'their' entailments turn out to be false does not even begin to make sense.

5 Against the anti-Austinian view, Part III

I will now give an independent argument in favour of the Austinian claim, which, if correct, undermines the anti-Austinian view at an even earlier juncture than the previous ones, since it focuses simply on the thesis that explicit performatives are truth-evaluable utterances rather than on the more specific thesis that they

are truth-evaluable utterances that are rendered true by being issued. The argument is based on the simple idea that if explicit performative utterances were truth-evaluable, they should be able, while retaining their performativity, to participate as premises in deductive arguments that would be recognised as instances of valid argument forms. For, since recognising such an argument as an instance of such a form requires only that its premises be recognised as truth-evaluable, rather than as true, nothing more—and, of course, nothing less—than its truth-evaluability would be required of any explicit performative utterance in order for it to be able to participate as a premise in a deductive argument that would be recognised as an instance of a valid argument form. So, an explicit performative's *in*ability to participate as a premise in an argument of that kind would constitute evidence that its performativity is incompatible with its truth-evaluability.

That performativity and truth-evaluability are indeed incompatible for this reason can be seen by considering (among several other similar cases) the sharp contrast in perceived validity between the 'argument' in (22) and the argument in (23),

(22) ! If you don't have a gun, I don't guarantee your safety.
        I guarantee your safety.
        Therefore, you have a gun.

(23)   If she didn't have a gun, he didn't guarantee her safety.
        He guaranteed her safety.
        Therefore, she had a gun.

or the equally sharp contrast in perceived validity between the 'argument' in (24) and the argument in (25):

(24) ! If you don't like pepper, I don't recommend to you the *poulet au poivre*.
        I recommend to you the *poulet au poivre*.
        Therefore, you like pepper.

(25)   If she didn't like pepper, he didn't recommend to her the *poulet au poivre*.
        He did recommend to her the *poulet au poivre*.
        Therefore, she did like pepper.

No one, I suppose, would claim that the 'arguments' in (22) and (24) are in any sense valid arguments if their respective second premises are read as explicit performative utterances (that is, if the second premise of (22) is read as *being* the guarantee that it names, and the second premise of (24) is read as *being* the recommendation that it names). On the other hand, everyone would accept that the arguments in (23) and (25), where the second premises *cannot* be read as explicit performative utterances, are classically valid arguments. But these sharply contrasting judgments about validity cannot be squared with the hypothesis that the second premises of (22) and (24) are *truth-evaluable* when they are read as explicit performative utterances. For, if the second premises of (22) and (24) *were* truth-evaluable when read as explicit performative utterances, it would be arbitrary in the extreme, and disrespectful of every compositionality requirement, *not* to formalise (22) and (24) as instances of the *same* argument schema that the formalisations of (23) and (25) would be instances of. And since that schema is classically *valid*, one would be forced by classical logic to take the patently *in*valid 'arguments' in (22) and (24) to be exactly on a par, in respect of their validity, with the clearly valid ones in (23) and (25). As a result, the only way to reconcile the parallel formalisations with the sharply contrasting judgments about validity would be to reject classical logic—in particular, to deny that *modus tollens* is an unrestrictedly reliable inference rule.

And, of course, if *modus tollens* were to be given up on such grounds, it should not be surprising that *modus ponens* would have to be given up as well. No one, presumably, would accept that the 'arguments' in (26) and (27) below are valid arguments, if their second premises are read as explicit performative utterances (that is, if the second premise of (26) is read as *constituting* the request that it names, and the second premise of (27) is read as *constituting* the promise that it names):

(26)　！　If I ask you to have dinner with me, you pretend you are busy.
　　　　　I ask you to have dinner with me.
　　　　　Therefore, you pretend you are busy.

(27)　！　If I promise to keep you safe, you don't trust me.
　　　　　I promise to keep you safe.
　　　　　Therefore, you don't trust me.

On the other hand, everyone would accept that the arguments in (28) and (29), where the second premises *cannot* be read as an explicit performative utterances, are classically valid arguments:

(28)   If he asks me to have dinner with him, I pretend I am busy.
        He asks me to have dinner with him.
        Therefore, I pretend I am busy.

(29)   If he promises to keep me safe, I don't trust him.
        He promises to keep me safe.
        Therefore, I don't trust him.

But this sharp contrast in perceived validity cannot be squared with the hypothesis that the second premises of (26) and (27) are *truth-evaluable* when read as an explicit performative utterances; for, if the second premises of (26) and (27) *were* truth-evaluable when so read, it would be compositionally irresponsible *not* to formalise (26)and (27) as instances of the *same* argument schema that the formalisations of (28) and (29) would be instances of; and since that schema is classically valid, one would be forced by classical logic to conclude that the obviously *in*valid 'arguments' in (26) and (27) are exactly on a par, in respect of their validity, with the obviously valid ones in (28) and (29). Consequently, the only way to reconcile the parallel formalisations with the sharp contrasts in perceived validity would again be to reject classical logic, concluding this time that *modus ponens*, no less than *modus tollens*, is an unreliable inference rule.

But rejecting classical logic would be quite an exorbitant price to pay in order to save the hypothesis that explicit performatives are truth-evaluable. And I am reasonably confident that those who have been attracted by that hypothesis would be quite unwilling to pay that price, and not pleasantly surprised to learn that the defence of their hypothesis would require them to become wanton logical revisionists. I therefore conclude that it would be in their best interests to reject that hypothesis and to accept that explicit performatives are *not* truth-evaluable, just as Austin, without offering arguments, has been urging.

6 Conclusion

Austin, as I mentioned in the beginning, claimed that his thesis that explicit performatives are not true or false is too obviously true to require argument: "It needs argument", he wrote (1975: 6), "no more than that 'damn' is not true or false." It is clear in hindsight that, in making that claim, Austin was greatly overestimating the degree to which subsequent philosophers would be prepared to accept without argument the view that explicit performatives are truth-valueless.

I have tried to show, however, that Austin's opponents can be given some good reasons for accepting that view, and that their best argument for not accepting it (namely, that by rejecting it they could explain how explicit performatives manage to accomplish the acts they name) turns out to be fatally flawed. If I am right, it is perhaps time to start suspecting that Austin's opponents may have been just too hasty in supposing that they have successfully neutralised the serious threat that explicit performatives pose to the bundle of prejudices that Austin was referring to as "the descriptive fallacy".

Chapter 5
Speaking of truth-telling: The view from *wh*-complements

1 The K-V thesis

Lauri Karttunen (1978) and Zeno Vendler (1980) have independently put forward the thesis (hereafter referred to as the "K-V thesis") that when one uses the verb *tell* with *wh*-complements one is logically committed to the view that the individual to whom one ascribes the act of telling is speaking the truth, whereas when one uses the verb *tell* with *that*-complements, one is not logically committed to the view that the individual to whom one ascribes the act of telling is speaking the truth.

Karttunen makes the point by saying, "The verb *tell* with a *that*-complement does not entail that what was told is true; with an indirect question it does" (Karttunen 1978: 172), and justifies that claim by citing the sentences *John told Mary that Bill and Susan passed the test* and *John told Mary who passed the test*, of which he says that the latter, unlike the former, "definitely says that John told the truth" (Karttunen 1978: 172). Vendler makes the same point by means of the following parable: "Consider the following two situations. Joe tells me that he lives in San Francisco. Later I visit him, and thus confirm that he indeed lives there. So I may report either that he told me that he lives in San Francisco, or, if I am more reticent, I may say that he told me where he lives. Jim also tells me that he lives in San Francisco. Later on, however, I find out that he actually lives in Oakland. Now in this case I still might report that he told me that he lives in San Francisco (and may add, if I so choose, that he lied). What I cannot do, however, is to report truthfully that he told me where he lives. For, in fact, he did *not* tell me where he lives. Thus, we have made a surprising discovery: telling *that* can be false, telling *what* cannot" (Vendler 1980: 283; cf. Vendler 1978).

I will argue below that the K-V thesis is not true; that its falsity has certain non-negligible implications on certain wider linguistic and philosophical issues; that a thesis that would be crucially different from it in that it would replace reference to entailments of truthfulness by reference to defeasible presumptions of truthfulness would be both accurate and capable of wider application; and that the reasons behind the truth of that *other* thesis could be sought in a theory of

linguistic conventions as solutions to coordination problems within groups of individuals. In the Appendix, I will show why an attempt to explain away the evidence against the K-V thesis fails, and will present some independently interesting evidence demonstrating its falsity. (Since it is doubtful whether the K-V thesis would make clear sense if it was taken to apply to infinitival *wh*-complements, I will throughout assume that it wasn't meant to so apply.)

2 Why the K-V thesis is not true

The simplest way of seeing why the K-V thesis is not true is to consider the following sentences:

(1)   John told his voters what he intended to do for them once elected, but, as usual, he was lying to them

(2)   John told his voters what he intended to do for them once elected, and, for once, he was not lying to them.

(3)   Old John told us whom he saw in the fog, but it turned out that he was mistaken. (The person he saw was Mr Smith, not Mr Brown.)

(4)   Old John told us whom he saw in the fog, and it turned out that he was not mistaken. (The person he saw was indeed Mr Brown.)

There is certainly no contradiction in either (1) or (3) and no redundancy in either (2) or (4). If the K-V thesis were true, however, both (1) and (3) would be contradictory—since their first conjuncts would entail that John told the truth whereas their second conjuncts would imply that he didn't tell the truth—and both (2) and (4) would be redundant—since their first conjuncts would entail that John told the truth whereas their second conjuncts would reaffirm that he indeed told the truth. So, the K-V thesis is simply not true. And this, it should be noted, is quite independent of the fact that the particular *wh*-complements occurring in the above sentences contain psychological predicates: The *wh*-complements of the following sentences do not contain any such predicates, and yet it is just as clear that neither of these sentences is contradictory, whereas the K-V thesis clearly entails that they both are.

(5)   John told us when he arrived in Paris, but the time he gave us cannot be correct: at that time he was still in Rome.

(6)   John told them where he has been between 4 and 5 p.m., but he was certainly lying since nobody was at the place he indicated during that time.

Given that the K-V thesis is clearly false, we may now turn to its consequences for Karttunen's and Vendler's more general theses.

3 What follows from the falsity of the K-V thesis

One of the main problems addressed in Karttunen (1978) was the problem of specifying the denotations of *wh*-complements (and, derivatively, of questions) in the framework of Montague semantics. Karttunen's solution to that problem was to let the denotations of *wh*-complements be sets of true propositions of a certain sort, and not, as in an earlier proposal within the Montague framework (Hamblin 1973) sets of true *or* false propositions of a certain sort. Karttunen's main argument in favour of his own solution (an argument that has never been questioned in the subsequent Montagovian literature on the topic) was that the verb *tell*, when it introduces *wh*-complements, entails that what is told is true, and that, therefore, if false propositions were excluded from the sets denoted by *wh*-complements, one could state the semantics of *tell*—and hence (given certain other assumptions of his framework) the semantics of interrogative complements in general—in a much simpler way than would otherwise have been possible. We have just seen, however, that the verb *tell* no more entails the truth of its *wh*-complements than it entails the truth of its *that*-complements. The conclusion must be, then, that Karttunen's argument has not been successful, and that a Montague semantics for interrogatives that is not committed to such an entailment is superior rather than inferior to one that is so committed. (Munsat [1986] is an otherwise very perceptive discussion of Karttunen and of *wh*-complementation in general; he explicitly endorses, however, the mistaken assumption that the K-V thesis is true.)

    Vendler's interest in the semantics of *tell* arose in connection with his thesis (first expounded in Vendler 1972) that thinking and speaking are activities directed toward objects of the same kind, and that their similarities in this respect can be ascertained by noticing that semantic distinctions observed in our talk about speech contents reflect semantic distinctions observed in our talk about mental contents. A fundamental such distinction in the mental domain, Vendler argues, is the one obtaining between objects of knowledge, which necessarily have the property of being true, and objects of belief, which do not necessarily have this property. But this same distinction, Vendler contends, obtains in the linguistic domain between objects of acts of telling that are expressed by *wh*-complements and objects of acts of telling that are expressed by *that*-complements:

"telling *that* can be false, telling *what* cannot" (Vendler 1980: 283). Furthermore, not only are the two domains, in this instance, exactly parallel, but there is, according to Vendler (1980: 288) dependence of the one on the other: When the content of an act of telling is reported by means of a *wh*-complement, it logically follows from the report that what the subject of the telling did was to "convey knowledge"; when, however, the content of an act of telling is reported by means of a *that*-complement, no such implication is present, and the subject of the telling is, at most, credited with a successful attempt to produce "mere belief". It therefore appears, Vendler concludes, that the semantics of *tell* provides a powerful argument in favour of the (for him, characteristically rationalist) thesis that the structure of the contents of speech is a reflection of the structure of corresponding mental contents. As we have seen, however, acts of telling whose contents are specified by means of *wh*-complements are no less compatible with the *falsity* of those contents than acts of telling whose contents are specified by means of *that*-complements. Consequently, Vendler's argument is unsuccessful, and, to the extent that his general thesis depends on its success, the prospects of the general thesis itself appear quite dubious. (Welbourne [1989] proposes an alternative philosophical exploitation of Vendler's data; his analysis also proceeds from the mistaken assumption that the K-V thesis is true.)

4 The truth behind the K-V thesis

Since there is no lack of evidence suggesting that failed entailment rules may sometimes be successful, albeit disguised, default rules (see, for example, Jackendoff 1983), it would be unwise to entirely dismiss the intuitions behind the K-V thesis without examining the possibility that what has been really triggering them is not anything as strong as the proposed entailment rule, but rather a *default* rule requiring that, *unless they contain explicit indications to the contrary,* ascriptions of telling are to be presumed to carry a commitment to the truth, rather than to the falsity, of what is reported as having been told. It seems, in fact, that such a rule would be worth positing. Consider the texts in (17) and (18):

(17) ? Mary never dated John. She was just telling us what they did on their first date and what they like doing since then.
(18) ? Mary doesn't have a son. She was just telling the journalists how old her son is and what he does for a living.

The oddity of these texts is undeniable, and it is not unreasonable to suppose that what provokes it is the fact that, given a sentence where someone is reported as having told something to someone else, there is a tendency to infer that the person doing the reporting believes that what has been said by the person being reported is true rather than false. This inference, however, is not logically compulsory, as can be seen from the fact that, when the reliability of the person being reported is openly denied by the reporter, the resulting text is a coherent and not an incoherent one:

(19) Mary never dated John. She was just telling us what they did on their first date, and what they like doing since then—but everyone who knows Mary can understand that she was lying.

(20) Mary doesn't have a son. She was just telling the journalists how old her son is and what he does for a living—but anyone who knows anything about Mary knows that she was not telling the truth.

What accounts for the contrast in acceptability between (17) and (18), on the one hand, and (19) and (20), on the other, cannot, therefore, be an entailment rule, but must be a default rule requiring that, *unless they contain explicit indications to the contrary,* ascriptions of telling are to be presumed to carry a commitment to the truth, rather than to the falsity, of what is reported as having been told. And that rule, apart for accounting for the evidence that shows that the K-V thesis mistakes a weaker semantic relationship for a stronger one, would also account for an important fact that the K-V thesis entirely misses—namely, that a defeasible presumption of truthfulness, rather than an entailment of truthfulness, is associated with the verb *tell* not only when it introduces *wh*-complements *but also* when it introduces *that*-complements. To see this, notice that (21), in which *tell* introduces *that*-complements, is just as odd as (22), in which it introduces *wh*-complements:

(21) ? Mary never met Harry. She was just telling us that he visits her every Monday, and that they play cards together.

(22) ? Mary never met Harry. She was just telling us how often he visits her and what game they play together.

However, just as it would be wrong to attribute the unacceptability of (22) to a supposed entailment from the telling ascriptions it contains to the truthfulness of the subject of those ascriptions, it would be wrong to attribute the unacceptability of (21) to a supposed entailment from the telling ascriptions *it* contains to

the truthfulness of the subject of those ascriptions. For, if the entailments in question existed, the truthfulness inferences would not be coherently deniable in either case, whereas in fact they are perfectly coherently deniable in both cases, as (23) and (24) show:

(23)   Mary never met Harry. She was just telling us that he visits her every Monday, and that they play cards together—but since Harry is long ago dead, and since he had never in his life played card games, Mary is either lying or simply confused about her visitors' identities.

(24)   Mary never met Harry. She was just telling us how often he visits her and what game they play together—but since Harry is long ago dead, and since he had never in his life played a game, Mary is either lying or simply confused about her visitors' identities.

Consequently, the proper way to account for the contrast between the unacceptability of either (21) or (22) and the acceptability of either (23) or (24) would be to posit not to an entailment rule, but rather a default rule requiring that an ascription of telling, whether it employs a *wh*-complement or a *that*-complement, is to be presumed to carry, unless it contains explicit indications to the contrary, a commitment to the truth, rather than to the falsity, of what is reported as having been told. The moral, then, is not, as the K-V thesis supposes, that *wh*-complements of *tell* enforce an entailment of truthfulness whereas *that*-complements of *tell* do not enforce an entailment of truthfulness, but rather that, though *neither* type of complement enforces an entailment of truthfulness, *both* types of complement trigger a defeasible *presumption* of truthfulness when introduced by the verb *tell*.

5 Where does the truth behind the K-V thesis come from?

Exactly why the interpretation of *tell* happens to be constrained by the default rule that has just been described is, of course, a different question, and one that some might prefer to leave unanswered. I want to conclude, however, by briefly noting that an interesting type of answer to that question might be attempted in the context of David Lewis's well-known proposal to explain the emergence of linguistic conventions as solutions to coordination problems within groups of individuals (Lewis 1969, 1975). Conventions in general arise, according to Lewis,

when several individuals have to coordinate their behaviour in order to satisfy an interest that they share; and conventions succeed in solving such coordination problems by virtue of being regularities in behaviour which are such that they are sustained in a population because (a) each member of the population expects all the others members of the population to conform to them, and (b) each member of the population conforms to them only because it expects that the other members will conform to them too (that is, would have no reason to conform to them if it did not expect that the other members will conform to them). Now, a *possible* language, according to Lewis, is simply a function from sentences to truth-conditions, and the task of a theory of *linguistic* convention it to explain under what conditions a merely *possible* language in that sense becomes the *actual* language of a human population. A possible language *L* becomes, on Lewis's theory of convention, the actual language of a human population *P* only if, thanks to an interest in communication shared by members of *P*, the following two behavioural regularities develop and come to prevail within *P*: first, each member of *P* avoids uttering sentences of *L* that he believes to be untrue, and does this *because* he expects that the other members will *themselves* avoid uttering sentences of *L* that *they* believe to be untrue (this is what Lewis calls a "Convention of Truthfulness in *L*"); and secondly, each member of *P* reacts to other members' utterances of sentences of *L* by believing that those sentences are true, and does this *because* he expects that the other members will *themselves* react to *his own* utterances of sentences of *L* by believing that those sentences are true (this is what Lewis calls a "Convention of Trust in *L*").

It is clear that if natural languages (or, more precisely, the truth-evaluable fragments thereof) are conventional communication systems in the sense of Lewis's theory of convention, then we should not be surprised to find that verbs that speakers of natural languages use in reporting other speakers' sayings may tend, in the absence of indications to the contrary, to be interpreted as conveying the reporting speaker's belief in the truth, rather than in the falsity, of those sayings—in other words, the reporting speaker's *trust* in other speakers' sayings. And this, as we have already seen, is precisely what happens with the verb *tell*, both when it introduces *wh*-complements and when it introduces *that*-complements. If, therefore, Lewis's proposal could be supposed to be in the right direction, the default rule we associated with *tell* would acquire a deeper significance: it would not be a mere idiosyncrasy of this verb that it can convey the speaker's belief in the truth, rather than in the falsity, of what is reported as having been told; it would rather be a symptom of the fact that its users belong to a population where it is generally preferred to *trust,* rather than to mistrust, other people's words, *because* (and to the extent that) one expects that other people will do the same

when it comes to one's *own* words. And this, it seems to me, would be as important a result as one might hope to obtain by investigating a single word's interpretation.

Appendix: Why the evidence against the K-V thesis won't go away[1]

Richard Holton (1997) has made an attempt to save the K-V thesis from the main argument against it developed in this essay, by explaining away the evidence on which that argument is based. The purpose of this appendix is to show that Holton's attempt has been unsuccessful, and to use this occasion for presenting some independently interesting data demonstrating that the K-V thesis is false.

The K-V thesis, it will be recalled, is the thesis that sentences where the content of an act of telling is reported by a *wh*-complement *entail* that what is reported as having been told is true (whereas sentences where the content of an act of telling is reported by a *that*-complement do not entail that what is reported as having been told is true). The main argument against that thesis developed above has two parts. The first part is that, if sentences where the contents of acts of telling are reported by *wh*-complements entailed that what is reported as having been told is true, then sentences like (1) and (3), repeated below for convenience, would be contradictory, whereas in fact they are not:

(1)   John told his voters what he intended to do for them once elected, but, as usual, he was lying to them.

(3)   Old John told us whom he saw in the fog, but it turned out that he was mistaken. (The person he saw was Mr Smith, not Mr Brown.)

The second part is that, if sentences where the contents of acts of telling are reported by *wh*-complements entailed that what is reported as having been told is true, then sentences like (2) and (4) would be redundant, whereas in fact they are not:

---

1 The paper on which the main text of the present essay is based (Tsohatzidis 1993a) has given rise to a reply by Holton (1997), to which I responded in Tsohatzidis (1997). This appendix is based on that response.

(2)   John told his voters what he intended to do for them once elected, and, for once, he was not lying to them.

(4)   Old John told us whom he saw in the fog, and it turned out that he was not mistaken. (The person he saw was indeed Mr Brown.)

Holton simply omits reference to this second part of the argument, without seeming to realise that a separate response to that part would be necessary if the position he advocates were to be sustained; for his position is that sentences where the contents of acts of telling are reported by *wh*-complements *do* entail that that what is reported as having been told is true; and anyone advocating that position has to explain not only why no one perceives (1) and (3) to be contradictory, if in fact they are, but also why no one perceives (2) and (4) to be redundant, if in fact they are; therefore, even if Holton's proposed explanation of the first of these facts were successful (which, as we shall presently see, it certainly is not), his position would be undermined by his inability to offer an explanation of the second one.

Holton's proposed explanation of the fact that no one perceives sentences like (1) and (3) to be contradictory, even though, in his opinion, they really *are* contradictory, is that everyone uttering them becomes blinded to their contradictoriness by indulging in a hitherto unknown psychological process called "protagonist projection". Thanks to that extraordinary process, Holton assures us, it is only when we utter the *second* conjuncts of (1) or (3) that we speak for ourselves, as it were; when, on the contrary, we utter their *first* conjuncts, we do not speak for ourselves, but, having "projected" ourselves into the position of the unfortunate victims of the deceitful acts of telling mentioned therein, we describe the situation of their deception as *they* would have described it without knowing that it *is* a situation of deception, and not as *we* would describe it *after* finding out that it was a situation of deception. Thus, when we utter the first conjunct of (1) we completely project ourselves, according to Holton, into the position of the innocent voters who have been duped by a malicious politician, and we therefore use the *wh*-complements that *the voters* would have used, presuming in their innocence that the politician has been honest; in the second conjunct, however, we speak strictly for ourselves, and we divulge to anyone interested in the matter that the politician has in fact been dishonest and that the voters have been duped. Similarly, when we utter the first conjunct of (3) we completely project ourselves into our former selves, who mistakenly thought that old John correctly identified the man he saw in the fog, and therefore use the *wh*-complements that *our former selves* would have used when trusting old John's veridicality; in uttering the second conjunct, however, we are back to our current selves who know better, and

who can inform anyone concerned that old John misidentified the man he saw in the fog, and that our past reliance on his words was misplaced.

On the assumption that this account is meant seriously, the following comments would be appropriate. First, the account contains no explanation as to why speakers should indulge in this peculiar process of vicariously changing personalities between their utterances of the conjuncts, and no psychological evidence that they do indulge in any such process. Second, the account fails to address the problem that, even if *speakers* were in fact indulging in vicarious personality changes while uttering the conjunctions, there is absolutely no guarantee that their *hearers* would simultaneously indulge in exactly parallel vicarious personality changes while interpreting the conjunctions, and therefore no reason to suppose that it is the occurrence of such putative changes that accounts for the acceptability judgments on the relevant sentences, which are *the same* for both speakers and hearers: even if *I* indulge, while uttering (1) or (3), in the "protagonist projections" that Holton imagines, it is certainly not necessary that *my hearers* should indulge in exactly analogous "protagonist projections" (or indeed in any "projections" whatsoever) while interpreting (1) or (3); but if this so, then the putative "projections" performed just by myself are quite irrelevant to explaining why *both* I and my hearers perceive (l) and (3) not to be contradictory—which, of course, was the main fact that Holton needed to explain. Third—and most importantly—, there is clear linguistic evidence that the alleged process of "protagonist projection", even if it were assumed to be both psychologically real and relevant in elucidating other linguistic phenomena, does not in fact determine any aspect of either the production or the interpretation of sentences where *tell* introduces *wh*-complements. Such evidence can be obtained in two ways. First, by comparing the behaviour of *tell* with the behaviour of certain other *wh*-complement taking verbs which should behave just like *tell* behaves *if* the "protagonist projection" account were correct, but which do not in fact so behave. And second, by considering certain further facts about the behaviour of *tell* for which the "protagonist projection" account should be excluded on conceptual grounds alone.

Evidence of the first kind can be obtained by considering the marked differences in interpretation between *wh*-complements introduced by *tell* and *wh*-complements introduced by such verbs as *reveal* or *divulge*. On the position defended in the present essay, a sentence where a *wh*-complement is introduced by *tell* would be expected to differ radically in interpretation from a sentence where a *wh*-complement is introduced by *reveal* or by *divulge*, since, although there is no a priori reason to suppose that the subject of an act of telling is truthful, there *is* a priori reason to suppose that the subject of an act of revealing or of an act of

divulging is truthful: nothing could be both untrue and a revelation, or both untrue and a divulgence. On Holton's position, on the other hand, sentences where *tell* introduces a *wh*-complement should be, in the respects that are relevant to the present discussion, semantically *exactly* like sentences where *reveal* or *divulge* introduce *wh*-complements, because the former *entail*, in his opinion, that what is reported as having been told is true, just as strongly as the latter entail, on anyone's opinion, that what is reported as having been revealed or as having been divulged is true. Contrary to what Holton should expect, however, the members of sentence pairs like those presented below differ quite sharply in acceptability, and the "protagonist projection" account is entirely incapable of explaining the contrasts between them:

(25) John told Helen where he was last night, but he was certainly lying, since last night he was definitely not where he told her he was.

(26) ? John revealed to Helen where he was last night, but he was certainly lying, since last night he was definitely not where he revealed to her he was.

(27) John told Mary where he was hiding, but it turned out that he was lying: He was not in fact hiding where he told her he was hiding.

(28) ? John divulged to Mary where he was hiding, but it turned out that he was lying: He was not in fact hiding where he divulged to her he was hiding.

According to Holton, the reason why we do not perceive (25) or (27) to be contradictory even though these sentences are, in his opinion, contradictory, is that we indulge in "protagonist projection" while uttering their first conjuncts. If that were the case, however, we should *also* be able to indulge in "protagonist projection" while uttering the first conjuncts of (26) and (28), and so end up believing that these sentences, too, are not contradictory, despite the fact that they really are. The fact, then, that we perceive *only* (26) and (28), and not (25) or (27), to be contradictory shows that Holton's explanation cannot be correct, and leaves as the only available explanation of the contrasts the one that is also the simplest: the reason why we perceive only (26) and (28), and not (25) or (27), to be contradictory is that only (26) and (28), and not (25) or (27), really *are* contradictory; and the reason for this is that, whereas sentences in which *reveal* and *divulge* introduce *wh*-complements do entail that the what has been revealed or has been divulged is true, sentences where the verb *tell* introduces a *wh*-complement do *not* entail that the what has been told is true.

We may now turn to evidence against the "protagonist projection" account that derives not from comparing the behaviour of *tell* with the behaviour of other *wh*-complement taking verbs, but by concentrating on certain especially interesting uses of *tell* itself, such as the following:

(29)   Both our agents told us where Peter died, but at least one of them must be mistaken, since the first told us that Peter died in Africa and the second that he died away from Africa.

(30)   The two clerks simultaneously told us when Peter was born, but at least one of them must be mistaken, since the first told us that Peter was born in 1956 and the second that he was born in 1965.

Just as would be predicted on the basis of the position defended in the present essay, and contrary to what would be predicted on the basis of the K-V thesis, neither of these sentences would be perceived as contradictory; and the simplest explanation of this fact is, of course, that neither of them *is* contradictory. Holton, however, is committed to treating each of them as contradictory, and must then claim that the reason why we are blinded to their alleged contradictoriness is that, when we utter their first conjuncts, we vicariously identify with our former gullible selves, whereas when we utter their second conjuncts we return to our current more enlightened selves. The problem, however, is that it is impossible, in these particular cases (and in several others of a similar kind), to suppose that, in uttering the sentences' first conjuncts, we could in fact vicariously identify with our former gullible selves, since the required identifications would lead us well beyond the limits of human gullibility: If I have in front of me two persons of which one tells me that Peter died in Africa and the other that he died away from Africa, it is impossible, no matter how gullible I am, that I should ever believe that they are *both* telling me the truth about Peter's place of death; but it is precisely the supposition that I *could* be misled into believing that they are *both* telling me the truth about Peter's place of death that would 'explain', on Holton's account, why, in uttering (29), I could without perceivable inconsistency both describe what these persons have been telling me and deny that they could both be truthful. Similarly, if I have in front of me two persons of which one tells me that Peter was born in 1956 and the other tells me that he was born in 1965, it is impossible, no matter how gullible I am, that I should ever believe that they are *both* telling me the truth about Peter's year of birth; but it is precisely the supposition that I *could* be misled into believing that they are *both* telling me the truth about Peter's year of birth that would 'explain', on Holton's account, why, in uttering (30), I could without perceivable inconsistency both describe what these persons

have been telling me and deny that they could both be truthful. In short: in order to be maintained in full generality, the "protagonist projection" hypothesis *must* attribute to speakers and hearers beliefs that no minimally rational person could have; and a hypothesis that can only be maintained at that cost cannot, of course, be accepted. As a final example in the same direction, consider the sentence in (31), which no one would perceive to be contradictory, even though, on the K-V thesis, it should be:

(31)   Yesterday in the asylum, Tom started telling us how he died and what he likes doing since his death—but, of course, Tom is not dead at all, he is just clinically certified as insane.

In order to explain, in conformity with the "protagonist projection" account, why no one would perceive that sentence to be contradictory, even though, on the K-V thesis, it should be, Holton would have to suppose that, in interpreting it, everyone temporarily becomes at least as insane as Tom is thereby claimed to be. For, on the "protagonist projection" account, the reason why we are blinded to the sentence's alleged contradictoriness is that, in interpreting its first conjunct, we identify with our former gullible selves who, in contrast to our current more enlightened selves, were perfectly willing to believe that Tom was telling the *truth* when telling us how he died and what he likes doing since his death. But since no minimally rational person could ever believe that someone could be telling the *truth* when telling others how he died and what he likes doing since his death, it follows that Holton's explanation can be maintained only on the supposition that speakers and hearers fail to be even minimally rational when interpreting each other's words; and an explanation that can only be maintained on that condition is not, of course, worth maintaining.

I conclude that Holton's attempt to save the K-V thesis through the "protagonist projection" fantasy has been unsuccessful—but also useful in providing some fresh opportunities for realising that the K-V thesis is simply not true.

Chapter 6
The distance between "here" and "where I am"

1 Introduction

Michael Dummett has claimed that there is an important distinction to be drawn
between two "features" of the meaning of a declarative sentence that he proposes
to call, respectively, "assertoric content" and "ingredient sense"; and he has con-
tended that certain widely shared philosophical views are erroneous by virtue of
failing to recognise that two declarative sentences may have the same "assertoric
content" while differing in "ingredient sense." Dummett has recently explicated
his proposed distinction as follows:

> What is the meaning of a declarative sentence? One answer might be: it is the
> principle that governs what it serves to convey to a hearer when the sentence
> is used on its own on any occasion to make an assertion, that is, how the
> hearer takes things to be if he accepts the assertion as correct. This is indeed
> an important feature of the meaning of the sentence, and it is how the ques-
> tion what the sentence means is often answered. But it is only one feature of
> the sentence's meaning: we may call it the *assertoric content* of the sentence.
> But it plainly does not constitute the whole of what the sentence means. We
> need to know, in addition, what contribution the sentence makes to the as-
> sertoric content of a more complex sentence of which it is a subsentence, and
> this is not in general determined by its own assertoric content. We may call
> this second feature the *ingredient sense* of the sentence. Two sentences may
> have the same assertoric content, but different ingredient senses. (Dummett
> 2004: 32)

My purpose in this essay is to argue that the example that Dummett has given as
representative of those that "may most easily be given" (Dummett 2004: 32) in
support of his proposed distinction—and that others, such as Robert Brandom
(Brandom 2009: 213), have accepted as unquestionably supporting it—does not,
in fact, support it.

Dummett's example—which is the only one that he uses to motivate the proposed distinction in his most recent exposition of it (Dummett 2010: 128–129)—is the following (the numbering of the quoted sentences is my own):

The two sentences

(1)   It is raining here.

and

(2)   It is raining where I am.

have the same assertoric content: if you believe a friend who, speaking to you on the telephone, utters either sentence, you learn exactly the same as if he had uttered the other sentence of the pair. But the sentences do not have the same ingredient senses, as is shown by the quite different meanings (assertoric contents) of the two sentences that result from inserting the quantifier "always":

(3)   It is always raining here.

and

(4)   It is always raining where I am.

The divergence occurs because the adverb "here" is temporally rigid, while the adverbial phrase "where I am" is temporally flexible. (Dummett 2004: 32–33)

There are, I shall argue, at least three problems suggesting that this example fails to provide evidence in favour of Dummett's proposed distinction between a sentence's "assertoric content" and its "ingredient sense."

2 Problem A

Dummett's assumption that, in sentences that are not subsentences of more complex sentences, the adverbial phrase "where I am" makes exactly the same contribution to "assertoric content" as the adverb "here", entails that there is no difference in "assertoric content" between the members of such pairs of simple sentences as (5a) and (5b), or (6a) and (6b):

(5)   a.   Winters are very heavy where I am.
         b.   Winters are very heavy here.
(6)   a.   It gets very humid where I am.
         b.   It gets very humid here.

If Dummett's assumption was correct, then, a speaker could not without contradiction assert a member of either pair while denying the other. It is clear, however, that a speaker *can* without contradiction assert a member of each pair while denying the other. For example, neither the sentence in (7) nor the sentence in (8) is contradictory, and each one of them could be truthfully uttered by a tourist who, in a conversation with inhabitants of the place he is visiting, compares the climate of his place of permanent residence with the climate of the place of his visit:

(7)   Where I am, winters are very heavy—unlike here.
(8)   Where I am, it gets very humid—unlike here.

And since, as (7) and (8) show, "where I am" is capable of making a contribution distinct from that of "here" to the "assertoric content" of even *simple* sentences, the fact that it is *also* capable of making a contribution distinct from that of "here" to the "assertoric content" of *complex* sentences (as Dummett's examples (3) and (4) were meant to demonstrate), shows that the evidence Dummett presents cannot support his proposed distinction between the "assertoric content" that a sentence has when it occurs on its own and the "ingredient sense" that it manifests when it occurs as a subsentence of a more complex sentence.

To circumvent the above argument, one might propose that the uses of "where I am" that are relevant to Dummett's discussion are not just any uses of "where I am", but *only* those in which "where I am" is paraphrasable by "where, *at this very moment*, I am". And one might point out, in support of that proposal,

that if, in (7) and (8), "where I am" is replaced by "where, at this very moment, I am" the resulting sentences—namely, (7′) and (8′)—are self-contradictory:

(7′)   Where, at this very moment, I am, winters are very heavy—unlike here.
(8′)   Where, at this very moment, I am, it gets very humid—unlike here.

The problem with this proposal, however, is that if the only uses of "where I am" that are relevant to Dummett's discussion are the ones in which it is paraphrasable by "where, at this very moment, I am" (which, incidentally, would mean that, contrary to what Dummett was supposing, the uses in question are temporally *rigid*, rather than temporally flexible), then Dummett's evidence fails to support his proposed distinction for a different reason—namely, because it now becomes impossible for the presence of the quantifier "always" to induce *any* relevant contrast, recognisable by Dummett, between simple and complex sentences: Just as there is no difference in "assertoric content" between the *simple* sentences (1) and (2′),

(1)   It is raining here.
(2′)   It is raining where, at this very moment, I am.

there is *also* no difference in "assertoric content" between the *complex* sentences (3) and (4′):

(3)   It is always raining here.
(4′)   It is always raining where, at this very moment, I am.

Rather than offering Dummett an escape route, then, the proposal under consideration forces upon him a dilemma: the interpretation of "where I am" that, on the proposal, would be required for the defence of his claim that his *simple* sentences are identical in "assertoric content" is precisely the interpretation under which he should abandon his claim that his *complex* sentences are *not* identical in "assertoric content"; and conversely, the interpretation of "where I am" under which he could maintain the latter claim is precisely the one under which he should abandon the former claim. It would not be unreasonable for someone to maintain, in view of this situation, that it is only by equivocating on the interpretation of "were I am" that Dummett was able to convince himself that this expression does not make a contribution distinct from that of "here" to the "assertoric

content" of simple sentences, but does make a contribution distinct from that of "here" to the "assertoric content" of complex sentences.

3 Problem B

Direct evidence against Dummett's assumption that "here" and "where I am" make exactly the same contribution to the "assertoric content" of the simple sentences where they occur is provided by the fact that the very sentences (1) and (2) that Dummett cites in support of that assumption manifest radically different interpretative profiles when placed in particular types of extra-linguistic and linguistic context. Regarding the effects of extra-linguistic context, two examples will suffice.

Example A. You are with a friend in your New York office, and want to convey to him the information, which you have just received on your mobile phone, that it is raining in Cairo. Pointing with your finger to Cairo on a map of Egypt hanging on one of your office walls, you could convey to your friend the information that it is raining in Cairo by saying to him,

(1)   It is raining here.

You could convey no such information, however, if you were to say to him,

(2)   It is raining where I am.

The reason, of course, is that, upon uttering (1) in the context under consideration, you would be construed as referring to a rainfall in Cairo rather than as referring to a rainfall in your New York office, whereas, upon uttering (2) in the same context, you would be construed as referring to a rainfall in your New York office rather than as referring to a rainfall in Cairo.

Example B. You occupy your assigned post in the open-air parking lot where you are currently employed, and it suddenly starts raining heavily. After a while, you are in your boss's office, asking for an umbrella. Your boss, who is quite unaware that a heavy rain has started, looks at you startled, and asks you why you need the umbrella. You could then explain to your boss why you need the umbrella by saying to him,

(2)   It is raining where I am.

You could provide no such explanation, however, if you were to say to him,

(1)   It is raining here.

The reason, of course, is that, upon uttering (2) in the context under consideration, you would be construed as referring to a rainfall in the open-air parking lot rather than to a rainfall in your boss's office, whereas, upon uttering (1) in the same context, you would be construed as referring to a rainfall in your boss's office rather than to a rainfall in the open-air parking lot.

Regarding the effects of linguistic context, it is sufficient to note that there are discourse environments where the use (2) is coherent because an anaphoric interpretation of "where I am" is possible when its deictic interpretation is blocked, whereas, in the same environments, the use of (1) is incoherent since no anaphoric interpretation of "here" is possible when its deictic interpretation is blocked. For example, there is a coherent interpretation of (2) when it occurs in a discourse such as (9), but there is no coherent interpretation of (1) when it occurs in a discourse such or (10):

(9)   Suppose a helicopter drops me in the middle of nowhere in a faraway country. It is raining where I am. What do I do?
(10)   ? Suppose a helicopter drops me in the middle of nowhere in a faraway country. It is raining here. What do I do?

And similarly, there is a coherent interpretation of (2) when it occurs in a discourse such as (11), but there is no coherent interpretation of (1) when it occurs in a discourse such or (12):

(11)   In the opening scene of the film you saw yesterday, I appear in the garden of a summer house on a winter night. It is raining where I am. What happens then?
(12)   ? In the opening scene of the film you saw yesterday, I appear in the garden of a summer house on a winter night. It is raining here. What happens then?

It is simply not true, then, that (1) and (2) have the same "assertoric content", if a sentence's "assertoric content" is taken to be what Dummett says it is—namely, "what [the sentence] serves to convey to a hearer when [it] is used on its own *on any occasion* to make an assertion" (Dummett 2004: 32; italics added). And if (1)

and (2) do not have the same "assertoric content" whenever they occur as free-standing sentences, the fact that, when they occur as subsentences of more complex sentences (as in Dummett's examples (3) and (4)), they may also have different effects on the "assertoric contents" of the sentences that are their hosts, cannot constitute evidence, as Dummett was supposing, in favour of his distinction between the "assertoric content" that a sentence has when it occurs on its own and the "ingredient sense" that it manifests when it occurs as a subsentence of a more complex sentence.

4 Problem C

Finally, telling evidence against Dummett's assumption that "here" and "where I am" make exactly the same contribution to the "assertoric content" of simple sentences in which they occur can be obtained by noticing that that assumption requires taking certain sentences that are unintelligible to be intelligible, and certain sentences that are not contradictory to be contradictory. Regarding the intelligibility-related cases, two examples will suffice.

Example A. Successively pointing with your finger to three different areas of your body, you could intelligibly say to your doctor:

(13)   I have pain here, here, and here.

However, you could not intelligibly say to him,

(14)   I have pain where I am, where I am, and where I am.

But (13) should have exactly the same "assertoric content" as (14), if Dummett's assumption was right. And since something unintelligible cannot, presumably, have the same "assertoric content" as something intelligible, it seems that Dummett's assumption was not right.

Example B. Opening your door to a person known to you to have been sincerely wondering about your whereabouts, you could intelligibly say to her:

(15)   Here is where I am.

You could not, however, intelligibly say to her:

(16)   Here is here.

(15) and (16), however, should have exactly the same "assertoric content", if Dummett's assumption was right. And since they do not, it is not.

Regarding the contradictoriness-related case, it is sufficient to note that, although sentences such as (17) or (18) are, on their normal interpretations, not contradictory,

(17)   Though he was indeed here, he was several yards away from where I am.
(18)   He was definitely here, but certainly not where I am.

sentences such as (19) or (20) are, on their normal interpretations, contradictory:

(19)   Though he was indeed here, he was several yards away from here.
(20)   He was definitely where I am, but certainly not where I am.

If Dummett's assumption was right, however, (17) and (18) should have exactly the same "assertoric contents" as (19) and (20), respectively. And since a contradictory sentence cannot, presumably, have the same "assertoric content" as a non-contradictory one, it is once more apparent that Dummett's assumption was not right.

5 Conclusion

Dummett has appealed to his proposed distinction between "assertoric content" and "ingredient sense" for important philosophical purposes—specifically, in order to argue that the widely accepted modal argument against descriptivist theories of names is defective (Dummett 1991: 48), and in order to argue that the even more widely accepted thesis that all (non-paradoxical) substitution instances of the schema ""p" is true if and only if p" are true is unwarranted (Dummett 2004: 36–37; 2007: 179–180). But these arguments would risk being regarded as unconvincing by his opponents *unless* they were supplemented by independent evidence in support of the distinction they appeal to—unless, that is, Dummett could produce cases, other than those under consideration in the particular disputes to which the arguments aim to contribute, that would be recognised by everyone as cases where identity of "assertoric content" coexists with non-identity of "ingredient sense." The example discussed in this essay was intended to provide precisely such independent evidence—indeed, the least controvertible such evidence, since it was supposed by Dummett to be representative of those that could

"most easily be given" in support of his proposed distinction. So, if I am right that the example *is* controvertible, and that it is controvertible for more than one reason, it seems that, though the possibility of finding independent evidence for Dummett's distinction cannot be excluded, the difficulty of finding such evidence has been significantly underestimated.[1]

Appendix: Dummett's subsidiary example

Besides his favourite example involving "here" and "where I am", which is the only one figuring in his most recent defence of the distinction between "assertoric content" and "ingredient sense", Dummett has given another example intended to motivate that distinction. Dummett's subsidiary example is the following (the numbering of the quoted sentences is again my own):

The sentences

(21)   I shall give you a D.

and

(22)   I intend to give you a D.

have the same assertoric content; but their ingredient senses differ, since the conditionals

(23)   If I give you a D, you will forfeit your grant.

and

(24)   If I intend to give you a D, you will forfeit your grant.

---

1 This conclusion is reinforced by the examination of another Dummettian example in the present essay's Appendix.

have different assertoric contents. (Dummett 2004: 33–34; Dummett adds, parenthetically, that "in English, the antecedent of the first conditional is grammatically in the present tense, but its sense is future.")

Curiously, Dummett does not make any attempt to justify the crucial assumption that he makes in presenting this example—the assumption, namely, that (21) and (22) have exactly the same "assertoric content"—, nor does he attempt to specify exactly what the allegedly identical content of (21) and (22) is supposed to be. There are, however, very good reasons for thinking that that crucial assumption is mistaken.

Notice that the simple sentences in (25) and (26) below clearly do *not* have the same assertoric content,

(25)   You shall give him a D.
(26)   You intend to give him a D.

and that, because of this, neither (27) nor (28) is in any sense contradictory:

(27)   You shall give him a D. That's not what you intend to do, but I'll kill you if you don't do it.
(28)   You intend to give him a D. But that's not what you shall do, since I'll kill you if you do it.

Given, then, that the simple sentences in (25) and (26) do not have the same assertoric content, one should assume that the simple sentences in (21) and (22) do not have the same assertoric content, either: it would be strange indeed if the obvious difference in content between (25) and (26) could be made to disappear *just* by switching the subject pronouns of these sentences from the second to the first person and their indirect object pronouns from the third to the second person.

Now, the obvious difference in content between (25) and (26) is the following: (25) is satisfied just in case its hearer goes on to actually *give* a D to a certain person (whether or not he currently *has the intentio*n of giving it), whereas (26) is satisfied just in case its hearer currently *has the intention* of giving a D to a certain person (whether or not he goes on to actually *give* it). But then, since semantic differences between verbs cannot be supposed to evaporate under changes in the grammatical person of the pronouns that surround those verbs, an exactly analogous difference should be assumed to exist between (21) and (22):

(21)   I shall give you a D.
(22)   I intend to give you a D.

(21) is *not* a statement about its speaker's intentions, and it is satisfied just in case its speaker goes on to actually *give* his hearer a D (whether or not he currently *has the intention* of giving it); (22), on the other hand, *is* a statement about its speaker's intentions, and it is satisfied just in case its speaker currently *has the intention* of giving his hearer a D (whether or not he goes on to actually *give* it).

Not surprisingly, the semantic difference between the first-personal examples (21) and (22) may take longer to notice than the exactly analogous semantic difference between the second-personal examples (25) and (26), since it tends to be obscured by the pragmatic fact that speakers' statements about their future behaviour are usually interpreted as evidence about their intentions, and speakers' statements about their intentions are usually treated as evidence about their likely behaviour. But this does not make the *contents* of (21) and (22) identical, and their difference in content becomes unmistakable when the pragmatic factors that tend to obscure it are explicitly filtered out. Thus, although (29) makes sense, (30) doesn't:

(29)   I do not intend to give you a D, but they will force me to. So, I shall give
         you a D.
(30)   ? I do not intend to give you a D, but they will force me to. So, I intend to
         give you a D.

And similarly, although (31) makes sense, (32) doesn't:

(31)   I intend to give you a D. Unfortunately, however, I will be dead before I
         manage to do that.
(32)   ? I shall give you a D. Unfortunately, however, I will be dead before I manage to do that.

Contrary to what Dummett assumes, then, there *is* a difference in "assertoric content" between (21) and (22) even when they occur as sentences that are *not* subsentences of more complex sentences. And if this is so, the fact that the complex sentences in which they may occur as subsentences can also differ in "assertoric content" (as Dummett's examples (23) and (24) were meant to demonstrate), cannot constitute evidence in favour of a distinction between the "assertoric content" that a sentence has when it occurs on its own and the "ingredient sense" that it can manifest when it occurs as a subsentence of a more complex sentence.

Consequently, Dummett's subsidiary example no more succeeds in motivating his proposed distinction than his favourite example involving "here" and "where I am" does. And since there is little beyond these two examples that Dummett offers as independent evidence in favour of his proposed distinction,[2] the overall case that he has made for it appears to be less than compelling.

---

2 The only other example that Dummett presents (all too briefly in Dummett 2004: 34) as independent evidence is, I believe, the weakest of the three, since its admission as evidence requires the unconditional acceptance of certain not widely shared views that he holds in the controversial area of the analysis of presupposition (in particular, the view that presupposition is a semantic rather than a pragmatic phenomenon).

Chapter 7
A problem for a logic of "because"

In a recent study entitled "A logic for *because*" (Schnieder 2011), Benjamin Schnieder observes that, although "the natural language connectives *and*, *or*, *not* and *if...then* are widely discussed" in philosophical logic, "considerably less attention has been paid to the explanatory connective *because*" (Schnieder 2011: 445). And he proposes to make a start at remedying that situation by developing a "logic for *because*" which is intended to make "aspects of our pretheoretic understanding of [that connective] explicit and precise" (Schnieder 2011: 447). The proposed logic aims to cover only "genuinely explanatory" uses of *because* (as opposed to "purely evidential" ones); and, among "genuinely explanatory" uses, it is intended to cover only those in which the explanation signalled by the presence of *because* is of a "noncausal" and not of a "causal" character (Schnieder 2011: 447).

The logic is presented as "an extension of a classical propositional natural deduction calculus" (Schnieder 2011: 448) and is claimed to be "consistent and moreover conservative" (Schnieder 2011: 452). Its centrepiece is a set of introduction rules for *because*, which are said to "exploit" the "core intuition" that a "sentence governed by a classical truth-functional connective has its truth-value *because of* the truth-values of the embedded sentences" (Schnieder 2011: 448).

The problem to be raised here about Schnieder's proposal can be exemplified by reference to several of his introduction rules, and a consideration of four among them will suffice to appreciate its nature. The rules in question (Schnieder 2011: 449) are the following:

(A) $$\frac{\varphi}{(\varphi \vee \psi) \text{ because } \varphi}$$

(B) $$\frac{\psi}{(\varphi \vee \psi) \text{ because } \psi}$$

(C) $$\frac{\psi}{(\varphi \rightarrow \psi) \text{ because } \psi}$$

(D) $$\frac{\neg\varphi}{(\varphi \rightarrow \psi) \text{ because } \neg\varphi}$$

The problem arises from the fact that natural language inferences such as those in (1)–(4) below would pretheoretically be regarded as *in*valid (if they could be interpreted at all), whereas—assuming the classical understanding of sentence connectives presupposed by Schnieder—related inferences such as those in (1′)–(4′) would be regarded as valid:[1]

(1)   Tom is alive. So, it is because Tom is alive that he is alive or dead.
(2)   London is in England. So, it is because London is in England that it is in Russia or in England.
(3)   John saw Mary. So, it is because John saw Mary that if he went home he saw Mary.
(4)   Fish don't speak Spanish. So, it is because fish don't speak Spanish that if fish speak Spanish Mary understands Kant.

(1′)   "Tom is alive" is true. So, it is because "Tom is alive" is true that "Tom is alive or Tom is dead" is true.
(2′)   "London is in England" is true. So, it is because "London is in England" is true that "London is in Russia or London is in England" is true.
(3′)   "John saw Mary" is true. So, it is because "John saw Mary" is true that "If John went home, John saw Mary" is true.
(4′)   "Fish don't speak Spanish" is true. So, it is because "Fish don't speak Spanish" is true that "If fish speak Spanish, Mary understands Kant" is true.

As can be verified by inspection, however, it is only the pretheoretically *in*valid inferences in (1)–(4), and *not* the valid ones in (1′)–(4′), that are substitution instances of the rules (A)–(D) of Schnieder's logic; indeed, there aren't *any* rules in Schnieder's logic by reference to which the validity of the inferences in (1′)–(4′) could be certified, despite the fact that these inferences faithfully represent the "core intuition" that was supposed to motivate the logic's construction. Schnieder's logic, therefore, faces the problem of, on the one hand, validating inferences that are not pretheoretically valid (such as those in (1)–(4)), and, on the other hand—and most importantly—failing to validate inferences (such as those in (1′)–(4′)) that not only appear to be valid but are also such that their perceived validity constitutes a manifestation of the very intuitions that the logic was meant to render "explicit and precise". Under these circumstances, it is unclear

---

1 In the interest of avoiding scope ambiguities that would otherwise arise in some of the examples, *because*-clauses have been uniformly clefted.

what exactly the logic's object should be supposed to be. In any case, it cannot be supposed to be the provision of a formal system exploiting Schnieder's guiding idea that a "sentence governed by a classical truth-functional connective has its truth-value *because of* the truth-values of the embedded sentences" (Schnieder 2011: 448), since inferences whose perceived validity is an expression of precisely that idea are simply not substitution instances of Schnieder's proposed formal rules.

The crucial difference between the valid inferences in (1′)–(4′) and the invalid ones in (1)–(4) is that, in the former, the connective *because* governs *explicit truth ascriptions* to certain sentences, whereas, in the latter, it governs *those sentences themselves*. It seems that attention to that difference, and to its consequences, is a prerequisite to the success of any attempt at constructing a logic of *because* of the sort that Schnieder was envisaging. It is worth noting, in this connection, that sentences of the form *That α explains why β*, which would be expected to be just as appropriate as *because*-sentences for expressing the sense of "noncausal explanation" that Schnieder is interested in elucidating, exhibit just the same sensitivity to whether *α* and *β* are replaced with *explicit truth ascriptions to certain sentences* or with *those sentences themselves*. When the former is the case, as in (5) and (6) below, the results make perfect sense:

(5)   That "London is in England" is true explains why "London is in Russia or London is in England" is true.

(6)   That "John saw Mary" is true explains why "If John went home, John saw Mary" is true.

When, however, the latter is the case, as in (7) and (8) below, the results hardly make sense:

(7)   That London is in England explains why it is in Russia or in England.

(8)   That John saw Mary explains why if John went home he saw Mary.

Indeed, it is only as causal pseudo-explanations of logical relations, rather than as non-causal explanations of logical relations, that (7) and (8) appear to understandable, if they are understandable at all.

Finally, it is worth noting that the markedly different behaviour of *because* in the two kinds of environment distinguished above (the one in which it governs explicit truth ascriptions to certain sentences and the one in which it governs those sentences themselves) is something that should be *expected* to arise, if it is

true—as Schnieder (2011: 453) explicitly admits that it is—that *because* is a hyper-intensional connective. For, as is well-known from discussions of disquotation-alism about truth, the clearest evidence for the hyperintensionality of *because* derives precisely from the fact that the sense, and the truth value, of a sentence where that connective occurs is highly sensitive to whether the ordered pair of sentences it joins consists of an explicit truth ascription to a sentence followed by that sentence, or, instead, of a sentence followed by an explicit truth ascription to that sentence. For example, the sentences in (9) and (10) below have evidently different senses, and one can consistently hold the first to be true while holding the second to be false, even though the only difference between them is that, in the first, *because* operates on the sentence pair ⟨*"Snow is white" is true*, *Snow is white*⟩, whereas, in the second, it operates on the sentence pair ⟨*Snow is white*, *"Snow is white" is true*⟩:

(9)   It is because snow is white that "Snow is white" is true.
(10)   It is because "Snow is white" is true that snow is white.

It should come as no surprise to anyone accepting the hyperintensionality of *because*, then, that a sentence where *because* governs explicit truth ascriptions to sentences and a sentence where it governs those sentences themselves, may have quite different logical properties. In particular, it should come as no surprise that a valid inference may be turned into an invalid one if, within that inference, a *because*-sentence of the former type is replaced with a *because*-sentence of the latter type. And this, it seems, is precisely what is happening when one moves from inferences such as (1′)–(4′) to inferences such as (1)–(4).[2]

---

2 Schnieder (2011) presents his "logic for *because*" as inspired by, and as consonant with, Kit Fine's work on the logic and metaphysics of grounding, but its problems are not necessarily problems for Fine's approach, since Fine does not intend his work as a contribution to the analysis of the natural language connective *because* (as Schnieder explicitly does), and does not regard *because* as wholly appropriate even for informal expositions of his theses. Thus, not only is the *because*-locution completely absent from Fine's formal logic of ground (Fine 2012a), but he elsewhere explicitly warns *against* reading the symbol for the grounding relation that he uses in his logic as equivalent to the ordinary word *because* (in Fine 2012b: 46, for example, he writes: "But, of course, *because* does not convey the distinctive sense of ground and is not able to distinguish between a single conjunctive antecedent and a plurality of non-conjunctive antecedents"). For a critical examination of Fine's work on grounding from a meta-metaphysical perspective, see Wilson (2014).

Chapter 8
What "lack" needs to have: A study in the semantics of privation

1 Introduction

George Lakoff (1987) has claimed that the verb *lack* offers a good opportunity for realising that lexical meaning is not a matter of relationships holding between words and mind-independent worlds (a view that he describes as characteristic of 'objectivist' semantics), but rather a matter of relationships holding between words and certain types of mental models that people use "for understanding the world and for creating theories about the world" (1987: 134)—a view that he regards (and recommends) as the central insight of 'cognitive' semantics. My purpose in this essay is twofold. First, to show that the analysis of *lack* that Lakoff offers in support of this claim is seriously defective and therefore not really serviceable in the context of the 'objectivism' vs. 'cognitivism' dispute. Second, to specify some elements of an alternative analysis, which, though far removed from what Lakoff regards as 'objectivist' assumptions, is equally far removed from the particular brand of 'cognitivism' that Lakoff is interested in promoting.

Lakoff's argument from *lack* (1987: 134–35) runs as follows. First, he claims that an 'objectivist' would regard a sentence like (1) as a paradigmatic instance of a sentence expressing an analytic truth,

(1)   Someone lacks something if and only if he doesn't have it.

and would therefore be committed to the view that *lack* and *not have* are paradigmatically synonymous expressions. Second, he claims that the fact that, in each of the following pairs of sentences, the second sentence is dubiously acceptable whereas the first it not, clearly shows that "*lack* and *not have* are not synonymous":

(2)   a.   My bike doesn't have a carburettor.
        b. ? My bike lacks a carburettor.

(3)   a   The Pope doesn't have a wife.
        b. ? The Pope lacks a wife.

Third, he claims that the nonsynonymy of *lack* and *not have* that these contrasts reveal is not only one that 'objectivists' have in fact overlooked but also one that they are in principle unable to acknowledge: in order for an 'objectivist' to count two expressions as nonsynonymous, the choice between them should make a difference to the states of the world represented by sentences containing them; but there is, Lakoff insists, not a single difference between the state of the world represented by a sentence containing *lack* and the state of the world represented by a corresponding sentence containing *not have*, the difference being rather that the sentence containing *lack* assumes, whereas the sentence containing *not have* does not assume, a particular folk theory concerning (not how the world actually is, but) how the world *should* be. Fourth, he claims that this distinctive role of *lack* could easily be captured by a non-'objectivist' definition that would make the presence of that word sensitive to the activation of "an idealised cognitive model with a background condition indicating that some person or thing *should have* something, and a foreground condition indicating that that person or thing does *not* have it" (Lakoff 1987: 134–135); and that this definition would clearly enable one to explain the oddity of sentences like (2b) and (3b) by appealing to the obvious fact that our folk theories about the world do not tell us that bikes should have carburettors or that the Pope should have wives, but, on the contrary, that "a bike should no more have a carburettor than the Pope should have a wife" (Lakoff 1987: 134–135). The case of *lack*, Lakoff therefore concludes, provides a good opportunity for recognising the virtues of 'cognitivist' and the vices of 'objectivist' approaches to the study of word meaning.

2 Amending the proposed analysis

Before going any further, it will be important to incorporate into the discussion a qualification without which both the 'cognitivist' and the 'objectivist' accounts of *lack* (as Lakoff presents them) would appear to be unworthy of serious attention. To see the need for that qualification, notice that, although it makes sense to say either

(4)   I should have this house alone/to myself.

or

(5)   I do not have this house alone/to myself.

it does not make any sense to say,

(6)   *   I lack this house alone/to myself.

And that, although it makes sense to say either

(7)   We should have this house together/in common.

or

(8)   We don't have this house together/in common.

it does not make any sense to say,

(9)   *   We lack this house together/in common.

It is clear that strings like (6) and (9) should be counted as perfectly meaningful both by Lakoff's 'objectivists' and by his 'cognitivists', since, on the view he associates with the former, (6) is synonymous with (5) while (9) is synonymous with (8), and, on the view he associates with the latter, (6) is synonymous with the conjunction of (4) and (5) while (9) is synonymous with the conjunction of (7) and (8). Since, however, it is equally clear that (6) and (9) are not meaningful, both the 'objectivist' and the 'cognitivist' views of *lack* would appear to have to be immediately rejected.

It seems to me that, in order to avoid this result, one should make a distinction between at least two senses of *have*—one in which it implies possession without necessarily implying ownership, and one in which it implies ownership without necessarily implying possession—and claim that it is the former, and not the latter, of these senses that *have* is intended to have both in the 'objectivist' and in the 'cognitivist' metalanguage. Since the examples that have just been cited as problematic for the 'objectivist' and 'cognitivist' accounts involve occurrences of *have* in which it is clearly paraphrasable by means of *own,* as in (10) and (11) below,

(10)   I do not own/I should own this house alone.
(11)   We do not own/We should own this house together.

and since, conversely, the occurrences of *have* in examples of the sort appearing in Lakoff's argument do not admit of such paraphrases—since, in other words,

(12) is no more a possible paraphrase of (2a) than (13) is a possible paraphrase of (3a)—

(12) * My bike doesn't own a carburettor.
(13) ? The Pope doesn't own a wife.

it seems that the proposed amendment is appropriate, and I will from now on assume that it has been adopted (without, of course, pretending to know how exactly the distinction between possession and ownership should ultimately be elucidated). The question before us, then, is the following: Can the problems apparently besetting the 'objectivist' definition of *lack* in terms of a single "not have" component (where *have* implies possession though not ownership) be made to disappear thanks to a 'cognitivist' definition supplying an additional "should have" component (where, again, *have* implies possession though not ownership)?

3 Why the proposed analysis fails

In order to demonstrate that the introduction of the proposed "should have" component would produce an incorrect definition of *lack,* one could, I presume, do either of two things: show that a statement where an entity is described as lacking a certain property does not lead to absurdity when conjoined with a statement denying that the entity in question should have that property; or show that a statement where an entity is described as lacking a certain property does lead to absurdity when interpreted as implying that the entity in question should have that property. Unfortunately for Lakoff's proposal, both of these things can be demonstrated quite straightforwardly.

Consider first the following sentences:

(14)   My proposal has many weaknesses, but it fortunately lacks arrogance, and it is precisely that feature that, in my view, no proposal should ever have.

(15)   Your essay isn't very original, but at least it lacks dogmatism, and, so far as I am concerned, that's the feature that no essay should ever have.

Neither of these sentences would strike anyone as absurd, but, on Lakoff's account of *lack,* they are as absurd as a sentence could be, since their speakers first imply, by virtue of using the verb *lack*, that certain entities of certain sorts should

have certain properties, and then openly assert that no entities of these sorts should ever have those properties. The "should have" component introduced by that account is, therefore, a source of special disadvantages, rather than a source of special advantages, for definitions of *lack* that would happen to incorporate it.

Consider next the following sentences:

(16)   As its name suggests, the empty set is the set that lacks members.
(17)   An intransitive verb, on the other hand, is a verb that lacks objects.
(18)   Vacuous names, of course, lack referents.

A mathematician, a grammarian and a logician might well utter (16), (17), and (18), respectively, in the course of trying to explain what the empty set, or an intransitive verb, or a vacuous name, is. But they would certainly find it absurd to add (or to be understood as implying) that the empty set is the set that does not have but *should* have members, that intransitive verbs are verbs that do not have but *should* have objects, or that vacuous names are names that do not have but *should* have referents. And since, on Lakoff's account of *lack,* these suggestions not only are not absurd but actually constitute part of what (16), (17) and (18) quite literally *mean*, it is once more apparent that the presence of the "should have" component would increase the inadequacy rather than the adequacy of definitions of *lack* that would happen to incorporate it.

I conclude that no statement of the semantics of *lack* would be correct if it included reference to a "should have" component, and that if the apparent difficulties of the 'objectivist' definition of *lack* are to provide an opportunity for interesting 'cognitivist' claims, a different kind of approach to the matter will have to be used. I will now sketch an outline of one such approach.

4 Sketch of an alternative analysis

Consider the sentences in (19), (20) and (21):

(19)   Vietnam lacks a capital.
(20)   Mary lacks a boyfriend.
(21)   Computers lack emotions.

In uttering them, their speakers do seem to assert what would be asserted by utterances of sentences like (22), (23) and (24), respectively,

(22)   Vietnam doesn't have a capital.
(23)   Mary doesn't have a boyfriend.
(24)   Computers don't have emotions.

but this cannot be *just* what they purport to be doing, otherwise it would be impossible to explain—among other things—why sentences like (26), (28), and (30) are, though false, acceptable, whereas sentences like (25), (27), and (29) are definitely unacceptable:

(25) ? Vietnam lacks a capital—indeed, there are no such things as capitals.
(26)   Vietnam doesn't have a capital—indeed, there are no such things as capitals.
(27) ? Mary lacks a boyfriend—indeed, there are no such things as boyfriends.
(28)   Mary doesn't have a boyfriend—indeed, there are no such things as boyfriends.
(29) ? Computers lack emotions—indeed, there are no such things as emotions.
(30)   Computers don't have emotions—indeed, there are no such things as emotions.

The distinctive purpose that *lack*-sentences are designed to serve is, I submit, twofold: on the one hand, to suggest that the entities referred to by their subject terms may be viewed as belonging to certain not explicitly mentioned *categories*, and, on the other hand, to suggest that the *prototypical* members of those not explicitly mentioned categories do possess the properties that the entities in question are being described as not possessing; in short, their distinctive purpose is to suggest that, by virtue of not having the properties that they are said not to have, the entities referred to by their subject terms become atypical members of certain not explicitly mentioned categories. On this account, then, what the speaker of *Vietnam lacks a capital* suggests, in addition to saying that Vietnam doesn't have a capital, is that Vietnam belongs to a category of things (namely, the category of states) whose prototypical members do have capitals (hence the oddity of example (25), whose speaker purports to say not only that Vietnam lacks a capital, but that there is nothing else in the world that does have a capital). Similarly, what the speaker of *Mary lacks a boyfriend* suggests, in addition to saying that Mary doesn't have a boyfriend, is that Mary belongs to a category of things (the category of girls of a certain age) whose prototypical members do have boyfriends (hence the oddity of example (27), whose speaker purports to say not only that Mary lacks a boyfriend, but that there is nothing else in the universe that does have a boyfriend). Finally, what the speaker of *Computers lack emotions*

suggests, in addition to saying that computers do not have emotions, is that computers belong to a category of things (namely, the category of things capable of intelligent operations) whose prototypical members (namely, human beings) do have emotions (hence the oddity of example (29), whose speaker purports to say not only that computers lack emotions, but that there is nothing else in the universe that does have emotions).

It will now be instructive to consider how the account just sketched can satisfactorily deal with examples that have already been shown to be impossible to correctly analyse within Lakoff's framework, and also how it can offer a superior analysis even of those few examples whose treatment within that framework might seem to be unobjectionable.

Consider first examples (16), (17) and (18) (repeated below for convenience):

(16)   As its name suggests, the empty set is the set that lacks members.
(17)   An intransitive verb, on the other hand, is a verb that lacks objects.
(18)   Vacuous names, of course, lack referents.

Although the "should have" glosses of these examples would be absurd, the ones suggested by the present account are perfectly intelligible. What the speaker of (16) suggests, on this account, is that, though the empty set does not have (nor should have) members, it belongs to a category of things (namely, the category of sets) whose prototypical elements do have members. What the speaker of (17) suggests is that, though intransitive verbs do not have (nor should have) objects, they belong to a category of things (namely, the category of verbs) whose prototypical members do have objects. And what the speaker of (18) suggests is that, though vacuous names do not have (nor should have) referents, they belong to a category of things (namely, the category of names) whose prototypical members do have referents.

Consider next examples (14) and (15) (again repeated for convenience):

(14)   My proposal has many weaknesses, but it fortunately lacks arrogance, and it is precisely that feature that, in my view, no proposal should ever have.
(15)   Your essay isn't very original, but at least it lacks dogmatism, and, so far as I am concerned, that's the feature that no essay should ever have.

Although the fact that these sentences are acceptable makes the proposed "should have" condition on the definition of *lack* clearly untenable, their accept-

ability is hardly surprising on the present account. What the speaker of (14) suggests, on this account, is that the fact that the particular proposal under consideration does not show arrogance makes it an atypical member of the category of *unsatisfactory* proposals (whose prototypical, for him, members are characterised by arrogance)—hence the implication that the proposal under consideration is less than fully unsatisfactory after all. And what the speaker of (15) suggests is that the fact that the particular essay under consideration does not manifest dogmatism makes it an atypical member of the category of *unsatisfactory* essays (whose prototypical, for him, members are characterised by dogmatism)—hence the implication that the essay in question is less than fully unsatisfactory after all.

Consider finally—and for reasons that will soon become apparent—examples (2b) and (3b), Lakoff's primary pieces of evidence in favour in his proposed analysis of *lack:*

(2)   b. ?   My bike lacks a carburettor.
(3)   b. ?   The Pope lacks a wife.

Given that, as Lakoff puts it, "a bike should no more have a carburettor than the Pope should have a wife", his analysis does predict the acceptability pattern in these particular examples. And it is easy to see that, on the analysis presently sketched, the acceptability pattern would be accountable as well: the oddity of (2b) would be attributed to the difficulty in imagining a situation in which the fact that a particular bike doesn't have a carburettor would make it an atypical bike (given that prototypical bikes do not have carburettors either); and the oddity of (3b) would be attributed to the difficulty in imagining a situation in which the fact that the Pope doesn't have a wife would make him an atypical member of the Catholic clergy (since no members of the Catholic clergy have wives either). It would appear, then, that, as far as these particular examples are concerned, one would be free to choose either account. In fact, however, the situation is more complicated, and suggests that even here it is the account presently proposed that should be retained.

Notice that since, on the present account, *lack*-sentences do not explicitly commit their speakers (or their interpreters) to a particular category with respect to which the typicality or nontypicality of the denotata of their subject terms is to be assessed, and since any given denotatum may obviously be categorised (and subsequently assessed for typicality or nontypicality) in many different ways, it is quite possible, within certain very broad limits, for the interpretative reactions to the same *lack*-sentence to differ widely, depending on the contextual avail-

ability of different categorisation clues. In particular, it is possible, on the account here proposed, for a given *lack*-sentence to be judged definitely unacceptable in a context where the only obvious categorisation of the denotatum of its subject term makes it difficult to regard that denotatum as atypical, and for the same *lack*-sentence to be judged definitely acceptable in a context where a different categorisation possibility has emerged, within which the denotatum of its subject term can easily be regarded as atypical. Now, Lakoff's account has absolutely no way of dealing with sharp variations of this sort in the acceptability of *lack*-sentences, and, as long as such variations remain a merely theoretical possibility, this might be taken not to matter. But they are not a merely theoretical possibility, as can be shown by reference to the very examples on which Lakoff has chosen to base his account: (2b) and (3b) can certainly be definitely unacceptable in many contexts, but they are perfectly acceptable in the context of the following dialogues:

(31)   *A*. No vehicle without a carburettor could ever do such a thing.
         B. My bike lacks a carburettor, but it did it.

(32)   *A*. Why on earth wouldn't you like to be the Pope?
         B. Well, there are many reasons. For one thing, the Pope lacks a wife.

The participants to these dialogues may know full well that "a bike should no more have a carburettor than the Pope should have a wife", but this, contrary to Lakoff's expectations, hardly prevents them from being intelligible to each other. The reason is that their dialogues offer them enough material for proceeding to other than the usual categorisations of the entities denoted by the subject terms of their *lack*-sentences. If a particular bike is thought of simply as a member of the category of bikes, then the fact that it doesn't have a carburettor could hardly make it an atypical member of the category of bikes, and the relevant *lack*-sentence will accordingly be felt to be odd. If, however, that particular bike is thought of as a member of the category of vehicles that have accomplished a certain spectacular feat, then the fact that it doesn't have a carburettor might well make it an atypical member of *that* category, and the very same *lack*-sentence will accordingly be felt not to be odd—which is exactly what happens with the *lack*-sentence in (31). Similarly, if someone thinks of the Pope merely as a member of the Catholic clergy, then the fact that he doesn't have a wife will hardly seem to make him an atypical member of the Catholic clergy, and the relevant *lack*-sentence will be felt to be odd. If, however, someone thinks of the Pope as a member of the category of lucky occupants of prestigious positions, then the fact that his

otherwise enviable position requires him not to have a wife may well be thought to make him an atypical member of *that* category, and the relevant *lack*-sentence will accordingly be felt not to be odd—which is exactly what happens with the *lack*-sentence in (32). In short, the presence of *lack* in a sentence of the form *X lacks Y* serves simply as an instruction to look for some category C in which X might plausibly be taken to belong, and which is such that the fact that X does not have the property Y would make it plausible to regard X as an atypical member of C. In the rare cases in which the search for such a category C will prove to be in vain, the *lack*-sentence will be found to be uninterpretable. But in the many cases in which that search will prove not to be in vain, the *lack*-sentence will have as many interpretations as there are contextually defensible allocations of things to categories.

5 Conclusions

The fact that human categorisation is, in most of its instances, an essentially context-dependent (and, hence, inherently unstable) phenomenon—in particular, that it depends not so much on retrieving, and then applying to particular situations, invariant representations of concepts from long-term memory, but rather on constructing variable representations of them from elements available in working memory—is one that is increasingly recognised in cognitive psychology (see especially Barsalou 1983, 1987; Barsalou and Medin 1986; Medin and Barsalou 1987), and it would be strange if linguistic categorisation turned out to be exceptional in this regard. The fact, then, that, on the proposed account, reference to contextually determinable rather than to antecedently determined, categories is a built-in semantic feature of *lack,* should hardly be viewed as surprising. But, though not really surprising, it has a potentially surprising (and so, noteworthy) implication on certain widespread views concerning the organisation of semantic description.

On these views (succinctly summarised in somewhat different terms by Atlas [1989: 25–65]), an adequate semantic description of a sentence should satisfy at least two demands. First, it should specify an interpretation that is context-invariant. And second, it should specify an interpretation that is fully determinate. Sentences containing indexical expressions cannot, of course, satisfy either demand, but this is usually regarded as a theoretically manageable problem, merely requiring a slightly more careful reformulation of the demands. On that reformulation, the first demand is that an adequate semantic description of a sentence should specify an interpretation that is context-invariant once the references of

the sentence's indexical expressions have been fixed. And the second demand is that it should specify an interpretation that is fully determinate once the references of the sentence's indexical expressions have been fixed. Under these reformulations the demands are considered to be clearly satisfiable, and to supply, in a sense, the *raison d'être* of semantic description. But *lack*-sentences are among clear cases of sentences showing that the two demands cannot simultaneously be satisfied. Even assuming that the references of the indexical elements of a sentence of the form *X lacks Y* have been fixed, its interpretation cannot be both fully determinate and context-invariant. For, in order to be fully determinate, it would have to specify the category C to which its utterer purports to allocate the entity referred to by X, and with reference to which he purports to claim that the fact that X doesn't have Y makes X an atypical member of C. But the category C cannot in general be fully determined just by considering the sentence itself, or by perusing an imaginary "Great Book" in which every given entity would have been allocated once and for all to one or more categories, irrespective of particular contexts of thinking and speaking. It has to be determined, rather, by considering who the utterer of the *lacks*-sentence is, what the context of his utterance is, and what the probable character and direction of his thought can plausibly be supposed to be within that context. And so, the closer the interpretation of the sentence would come to being determinate, the further would it be removed from being context-invariant.

One of the implications of this last point that seems appropriate to mention in concluding concerns the different perspective that it offers on the 'objectivism' versus 'cognitivism' dispute referred to at the beginning of this essay. The objectivist's claim that, in uttering a sentence of the form *X lacks Y,* a speaker just means that X does not have Y, cannot, of course, be right, since it cannot explain why the acceptability ranges of *X lacks Y* and *X does not have Y* can diverge. It does seem true, however, that the only fully *determinate* thing that each speaker of a sentence of the form *X lacks Y* explicitly says is, indeed, that X does not have Y—the remaining (and more interesting) part of what he means being not fully determinate but *determinable* on the basis of contextual information. So, the objectivist's most important mistake was to produce a description capable of acknowledging only the determinate and not the merely determinable aspects of a sentence's sense. But it would be just as mistaken on the cognitivist's part to try to supply the elements missing from the objectivist account in determinate, rather than in merely determinable, form—and this is just what Lakoff's "idealised cognitive model" for *lack* tries, and fails, to do, by assuming that there is but one, context-invariant, answer to the question as to whether any given entity does or does not belong to any given category, and, hence, that there cannot be but one

right answer to the question as to whether any given entity "should" or "should not" have any given property. In an important sense, then, the 'objectivist' and the 'cognitivist' accounts of *lack* appearing in Lakoff's discussion suffer from a common mistake. And the real issue that emerges once that mistake is recognised is not whether mental categories play a role in semantic interpretation, but whether the role they undoubtedly do play in semantic interpretation is or is not best accounted for by conceiving of them as essentially contextual, and hence as inherently unstable, constructs. At least as far as the interpretation of *lack* is concerned, the answer to that question must, I think, be affirmative.

Chapter 9
A fake typicality constraint on asymmetric acceptability

1 Introduction

If things that have the property *p* are things that either necessarily have or necessarily fail to have the property *q*, then a semantic account of "*p*" should presumably alert us to this fact. If, however, things that have the property *p* are things that neither necessarily have nor necessarily fail to have the property *q*, but are, rather, things that *typically* have the property *q*, should or should not a semantic account of "*p*" keep us informed of this fact? To put it in a terminology that is sometimes preferred, should or should not a semantic account of "*p*" allow that our dictionary knowledge not always be strictly separable from our encyclopedic knowledge? Several semantic theorists who believe that this question should be answered in the affirmative have given a specific linguistic reason in favour of their belief—namely, that, unless it was answered in the affirmative, one could not explain certain quite clear acceptability and unacceptability judgements concerning compound sentences of the forms *S but S′* and *S but ~S′*, where *S* is a sentence in which a property *p* is predicated of an entity *t* and *S′* is a sentence in which a different property *q* is predicated of the same entity *t*. More specifically, the argument is that cases where the compound sentences in question are neither both acceptable nor both unacceptable can only be explained by supposing that the unacceptable among them frustrates, and that the acceptable among them does not frustrate, an expectation, induced by the use of *but*, concerning the typicality or non-typicality of the association or non-association of the property *q* with things that possess the property *p*.

   Consider—to take three examples that have been cited as representative—the following pairs of sentences:

(1) ? He's lost his watch, but he doesn't know where it is.
(2)   He's lost his watch, but he knows where it is.

(3) ? It's a dog, but it can bark.
(4)   It's a dog, but it can't bark.

(5) ? She is a mother, but she is a housewife.

(6)   She is a mother, but she isn't a housewife.

Bendix (1966: 24) suggests that the contrast in acceptability between (1) and (2) is due to the fact that the expectations of non-typicality raised by the use of *but* are frustrated in the case of (1), which ascribes to a person who has lost something a property that such persons typically have (namely, that of not knowing where the lost thing is), whereas they are not frustrated in the case of (2), which ascribes to a person who has lost something a property that such persons typically fail to have (namely, that of knowing where the lost thing is). Cruse (1986: 17) suggests that the contrast in acceptability between (3) and (4) is due to the fact that the expectations of non-typicality raised by the use of *but* are frustrated in the case of (3), which ascribes to a certain dog a property that dogs typically have (namely, that of being able to bark), whereas they are not frustrated in the case of (4), which ascribes to a certain dog a property that dogs typically do not have (namely, that of being unable to bark). And Lakoff (1987: 81) suggests that the contrast in acceptability between (5) and (6) is due to the fact that the expectations of non-typicality raised by the use of *but* are frustrated in the case of (5), which ascribes to a given mother a property that mothers typically have (namely, that of being housewives), whereas they are not frustrated in the case of (6), which ascribes to a given mother a property that mothers typically do not have (namely, that of not being housewives).[1]

I would like to suggest that the argument from *but* that these and many other such claims exemplify does not succeed in conferring linguistic significance on typicality considerations. I will first argue that there are clear typicality contrasts which do not correspond to any contrasts in acceptability within substitution instances of the two *but*-frames. I will then argue that there are clear contrasts in acceptability within substitution instances of the two *but*-frames to which there do not correspond any possible typicality contrasts. And I will finally suggest that even if our typicality and acceptability judgements corresponded to each other in

---

1 The terminologies these and other authors use for naming the semantic traits they extract by applying the *but*-test vary widely—Bendix, for example, calls them "connotative", Cruse calls them "expected", and Lakoff calls them sometimes "stereotypical" (1987: 81) and sometimes "prototypical" (1987: 404). Weinreich (1963: 194) was apparently the first author to recommend the use of *but*-sentences as detectors of semantically relevant information, but the way his mature theory (Weinreich 1972) would propose to deal with that information seems to be quite different from, if not opposed to, the way the above mentioned authors would envisage (cf. Lehrer 1990: 370–371).

the way the argument from *but* predicts, the theoretical conclusions that its proponents would like to derive on the basis of these correspondences about the inseparability of world-knowledge and word-meaning would not in fact follow from them.

2 The problem of sufficiency

That a contrast in typicality is not a sufficient condition for a contrast in acceptability within the *but*-frames can be appreciated by considering the following dialogues:

(7) *A*. Some people have asked me to do things without really wanting me to do them.
B. As for me, I did ask you to do certain things, but I really wanted you to do them.

(8) *A*. Most people who asked me to do things really wanted me to do them.
B. As for me, I did ask you to do certain things, but I didn't really want you to do them.

(9) *A*. Some people read a lot, but learn nothing.
B. I did read a lot, but I did learn something.

(10) *A*. Everyone who reads a lot ought to be learning something.
B. I did read a lot, but I learned nothing.

If there are any typicality statements that one should accept as true, then one should certainly accept as true the statement that people who ask other people to do things typically want, rather than do not want, them to do those things, or the statement that people who do a lot of reading are typically people who do not thereby fail to acquire any amount of learning. But if this is so, then the proponent of the argument from *but* should predict that only B's contributions to the dialogues in (8) and (10), and *not* B's contributions to the dialogues in (7) and (9), would be acceptable. The above dialogues show, however, that, just as there are contexts where the former contributions are indeed acceptable, there are other contexts where the latter are also acceptable. And since the difference between the two types of dialogue can hardly be traced to a difference in the relevant typicality assumptions that would need to be made in order to interpret them (since,

in other words, the interpretation of *all* the dialogues in question is consistent with the assumption that requests *typically* emanate from real rather than from feigned desires, and that extensive reading *typically* results in some rather than in no amount of learning), it follows that there can be quite clear contrasts in typicality to which there need not correspond any contrasts in acceptability within the two *but*-frames.

3 The problem of necessity

That a contrast in typicality is not even a necessary condition for a contrast in acceptability within the *but*-frames can be appreciated by considering the following pairs of sentences:

(11)   Your father is seriously ill, but he hasn't died.
(12)   ? Your father is seriously ill, but he has died.

(13)   I misled you into thinking that I am a Catholic, but I am not a Catholic.
(14)   ? I misled you into thinking that I am a Catholic, but I am a Catholic.

(15)   We have finished our job with some difficulty, but we have finished it.
(16)   ? We have finished our job with some difficulty, but we haven't finished it.

Since the contrasts in acceptability between the members of these pairs are indisputable, the proponent of the argument from *but* would have to appeal to certain typicality contrasts in order to account for them. But the typicality contrasts he would have to appeal to are certainly non-existent. In order to account for the oddity of (12) as opposed to the acceptability of (11), he would have to claim that it is typical of (though not necessary for) persons who are seriously ill to be, at the same time, persons who are dead. In order to account for the oddity of (14) as opposed to the acceptability of (13), he would have to claim that it is typical of (though not necessary for) people who induce in their audience the false belief that they are Catholics to actually be Catholics. And in order to account for the oddity of (16) as opposed to the acceptability of (15), he would have to claim that it is typical of (though not necessary for) persons who have finished their jobs with some difficulty to be persons who have not finished those jobs. These claims are, of course, not just contingently but *necessarily* false. And the reasons why they are false should have led the proponent of the argument from *but* to expect that *both* members of each of the pairs of sentences just cited will be unaccept-

able, the one on account of its contradictoriness and the other on account of its redundancy. (Cf. Cruse's (1986: 123) suggestion to the effect that the unacceptability of both *It's a horse, but it's an animal* and *It's a horse, but it isn't an animal* shows that being animals is a necessary, and not merely a typical, property of horses; or Bendix's (1966: 26–27) claim to the effect that the unacceptability of both *He lost it, but unintentionally* and *He lost it, but intentionally* shows that intentionality is an excluded, and not merely an atypical, property of events of losing.) The fact, then, that only *one* member of each of the sentence pairs previously cited happens to be unacceptable, and thus leads the proponent of the argument from *but* to mistake excluded traits for merely atypical ones and necessary traits for merely typical ones, shows that a genuine asymmetry in typicality is not even a necessary condition for an asymmetry in acceptability within the two *but*-frames.

4 A final problem

Given that typicality contrasts are neither necessary nor sufficient for the existence of acceptability contrasts within the *but*-frames, the hypothesis of a causal relation between these two types of contrast should obviously be rejected; I believe, however, that there is an independent, and important, reason for rejecting that hypothesis even when one confines one's attention to cases where one would be prepared to grant to the proponent of the argument from *but* both his assumptions about typicalities and his observations about acceptabilities. To see this, consider the following pairs of sentences:

(17) ? It's wine, but it tastes like wine.
(18)   It's wine, but it doesn't taste like wine.

(19) ? It's a violin, but it produces a violin-like sound.
(20)   It's a violin, but it doesn't produce a violin-like sound.

(21) ? It's a cat, but it behaves like a cat.
(22)   It's a cat, but it doesn't behave like a cat.

Let us assume, as the proponent of the argument from *but* would have done in attempting to deal with these examples, that there is a particular kind of taste that wine typically has, that there is a particular kind of sound that violins (when appropriately handled) typically produce, and that there is a particular kind of

behaviour that cats typically exhibit. It would then be reasonable to suppose that a person would be unable to be surprised at the fact that a particular wine does not have the typical wine-like taste, or that a particular violin does not produce the typical violin-like sound, or that a particular cat does not behave in the typically feline way, *unless* that person knew how wine typically tastes, or how violins typically sound, or what cats typically do. And since the proponent of the argument from *but* is committed to claiming that it is precisely our ability to actually distinguish typical from atypical instantiations of properties in the real world that determines our willingness or unwillingness to utter sentences obtainable from the two *but*-frames, he should expect to find that our capacity to judge some of the sentences just cited as appropriately utterable and some others as not appropriately utterable depends on our having experiences enabling us to verify the relevant typicality facts. But this is evidently not the case. I may have never tasted wine, but my knowledge of English leaves me in no doubt that (17) is an odd thing to say whereas (18) is not. You may have never heard a violin sound, but your knowledge of English should suffice to tell you that (19) is an odd thing to say whereas (20) is not. And a third person may know nothing about the behavioural habits of cats, and yet know enough English to be able to determine that (21) is an odd thing to say whereas (22) is not. It seems, therefore, that, contrary to the basic theoretical claim that the proponents of the argument from *but* were hoping to derive from it, the capacity to recognise acceptability contrasts within the *but*-frames is in principle independent of the capacity to make the corresponding typicality discriminations in the real world. And this conclusion emerges even more dramatically when one considers that there are acceptability contrasts exactly analogous to the ones just noticed, whose recognition not only need not rely but *could not possibly* rely on the capacity to make typicality discriminations in the real world. For example, the contrast in acceptability between (23) and (24), or between (25) and (26), is no less clear than the contrasts previously noticed,

(23)   ? It's a unicorn, but it behaves like a unicorn.
(24)   It's a unicorn, but it doesn't behave like a unicorn.

(25)   ? It's a round square, but it looks like one.
(26)   It's a round square, but it doesn't look like one.

and yet it would be nonsensical to suggest that the recognition of these contrasts in a linguistic population depends on the ability of its members to distinguish between typical and atypical unicorn behaviours, or between typical and atypical

round square shapes, since there exist *no* such things as unicorns and round squares to begin with.

5 Conclusion

The moral to be drawn, I suggest, is that the argument from *but* has failed—and, in view of the considerations of the last section, could not but have failed—to isolate a type of construction that would force one to acknowledge the necessity of recognising, alongside semantic traits that are absolutely required or absolutely prohibited, a special class of typicality traits that would occupy an intermediate position in semantic space. This does not mean, of course, that all typicality traits that have ever been accepted in the literature were detected by means of the two *but*-frames. Since, however, the argument's popularity suggests that many of them may have been so detected, it is legitimate to suspect that many of them may have been illegitimately accepted. [2]

Appendix: An unsuccessful defence of the argument from "but"

The original publication of the main text of the present essay (Tsohatzidis 1992a) has given rise to a reply by D. A. Cruse (1992), to which I responded in Tsohatzidis (1992b). The present Appendix reproduces that response.

Cruse (1992) begins by (though, fortunately, he does not persist in) misconstruing my position as directed against a claim to the effect that "appropriate use of *but* is always and everywhere determined by expected traits associated with individual lexical items" (Cruse 1992: 399). I am not aware of any such claim any more than Cruse is, and it is certainly not the one discussed in Tsohatzidis (1992a): both the claim I discuss and my counterexamples to it concern a small, and clearly circumscribed, subset of *but*-sentences—namely, those of the forms *S but S′* and *S but ~S′* in which *S* is a subject-predicate sentence where a certain property *p* is predicated of an entity *t* and *S′* is a subject-predicate sentence where a different property *q* is predicated of the same entity *t*. Since these are obviously

---

2 For some comments on an unsuccessful attempt to resist this conclusion, see the present essay's Appendix.

not the only kinds of *but*-sentences that exist, the suggestion that I might be purporting to discuss what *but* "always and everywhere" does would appear to be unduly impatient.

What is the content of the claim that I actually discuss? As I reconstruct it (and Cruse offers no argument against that reconstruction), the claim in question involves three theses, whose joint effect is supposed to be a demonstration of the importance of recognising, alongside semantic traits which reflect necessary connections or non-connections between properties, a special class of semantic traits reflecting typical, though not necessary, connections between properties. The first thesis is that if the necessary connection or non-connection between properties was the only semantically pertinent feature of the relation between words expressing those properties, there would be only two acceptability patterns for *but*-sentences (of the specified forms) to follow: either the connection (or lack thereof) would be contingent, in which case the relevant *but*-sentences would both be acceptable, as in (27) and (28) below,

(27)   It's a dog, but it's friendly.
(28)   It's a dog, but it isn't friendly.

or the connection (or lack thereof) would be necessary, in which case the relevant *but*-sentences would both be unacceptable (the one because of redundancy and the other because of contradiction), as in (29) and (30) below:

(29)   ? It's a dog, but it's an animal.
(30)   ? It's a dog, but it isn't an animal.

The second thesis is that there are, nevertheless, pairs of *but*-sentences (of the specified forms) which follow neither the one nor the other of these acceptability patterns—in other words, pairs of *but*-sentences of which one member is acceptable and the other member unacceptable, as in (3) and (4), repeated below for convenience:

(3)   ? It's a dog, but it can bark.
(4)   It's a dog, but it can't bark.

And the third thesis is that what *explains* the asymmetric acceptability pattern in all pairs of sentences of this last sort is that the properties therein described as being connected or non-connected are *typically*, though not necessarily, connected in the real world.

Clearly, it is the last of these three theses that, if true, would guarantee the significance of the view under discussion: if it could be shown that typicality considerations cannot be the ones that explain the phenomenon of asymmetric acceptability within the two *but*-frames, it would be futile to insist that observation of acceptability patterns within these frames is relevant in motivating claims about the admissibility or inadmissibility of typicality traits in semantic theory. And in order to show that typicality considerations *cannot* play that explanatory role, one could show either that the acceptability patterns can be asymmetric even in the absence of relevant typicality contrasts, or that the acceptability patterns can be symmetric even in the presence of relevant typicality contrasts. I believe that both of these things are clearly shown in Tsohatzidis (1992a), and I am therefore not surprised to find that Cruse's attempts to question my arguments turn out to be attempts to change the subject.

Consider first my claim that the acceptability can be symmetric even in the presence of relevant typicality contrasts. It is based on, among other things, the observation that the *but*-sentences in the examples below are both clearly acceptable, even though the logic of the argument from but would require only *one* of them to be acceptable, since the sentences provide ample room for invoking the relevant typicality contrast (since, in other words, it is typical, rather than atypical, of persons who make requests for actions to want, rather than not to want, those actions to be performed):

(7)   *A*. Some people have asked me to do things without really wanting me to do them.
        B. As for me, I did ask you to do certain things, but I really wanted you to do them.

(8)   *A*. Most people who asked me to do things really wanted me to do them.
        B. As for me, I did ask you to do certain things, but I didn't really want you to do them.

Cruse's response to this argument is, in effect, that, although both of the *but*-sentences are, in the contexts provided, acceptable, there are other contexts (affectionately called "straightforward") where only one of them appears to be acceptable. But it is hard to see how this could be supposed to save the thesis under discussion. If a test predicts that a certain sentence will be *un*acceptable, all that is required in order to falsify its prediction is to produce at least one natural context in which that sentence *is* acceptable. Since, as Cruse himself admits, such a context has been produced, the prediction has been falsified, and there is nothing

more that need or can be said on the matter. Of course, Cruse may have hopes of devising a new version of the *but*-test that would sidestep this problem by operating not with the familiar distinction between acceptability and unacceptability but with the hitherto unknown distinction between "acceptability in straightforward contexts" and "unacceptability in straightforward contexts". Until, however, the obscure (and, to my mind, mythical) distinction between "straightforward" and "non-straightforward" contexts is given a formulation that is both precise and non-question-begging, the only *but*-test that can be supposed to exist is the one whose predictions have actually been falsified.

Consider next my claim that the acceptability can be asymmetric even in the absence of relevant typicality contrasts. This is principally based on the observation that, of the two sentences below, the one is clearly acceptable and the other clearly unacceptable, even though the logic of the argument from but would require *both* of them to be equally unacceptable, since the one is redundant and the other contradictory (because it is not just 'uncommon' but quite impossible for persons who are seriously ill to be persons who are dead):

(11)　　Your father is seriously ill, but he hasn't died.
(12)　? 　Your father is seriously ill, but he has died.

Cruse concedes that there is a problem here, but misidentifies its origin. He thinks that the problem could be avoided by requiring that the only form of unacceptability relevant to the operation of the *but*-test should be that due to redundancy and not that due to contradiction. He appears to forget, however, that, even with this restriction (which, of course, does not belong to the original formulation of the test), there would be no way to avoid the important fact that the examples reveal—namely, that redundancy does not invariably lead to unacceptability, and that, although *Your father is seriously ill, but he hasn't died* is no less redundant than *It's a dog, but it's an animal* (since a dog is an animal just as necessarily as a man who is seriously ill is a man who is alive), it is *only* the latter, and not the former, of these sentences that is actually unacceptable. Now, if redundancy does not invariably lead to unacceptability even in *but*-sentences invoking necessary connections, there can surely be no reason for expecting it to lead, rather than not to lead, to unacceptability in *but*-sentences invoking merely typical connections. And since the *but*-test certainly embodies the assumption that such an expectation would *be* reasonable, the rationale behind its employment simply disappears. Of course, this might well prompt Cruse to devise a new kind of argument, which, though somehow involving reference to the *but*-frames, would entirely dispense with employing the notions of acceptability or unacceptability in

evaluating their substitution instances. However, no such argument has ever been proposed, and the one before us must be rejected for, among others, the reason just cited.

Cruse's reaction to my final objection to the argument from *but* gives me no reason to modify my view that it should be treated as a deeply disturbing objection. The point was that real world knowledge of relevant typicality contrasts should be truthfully ascribable to users of *but*-sentences if the proposed explanation of asymmetric acceptability within the *but*-frames was to have more than ritual significance, and that the requisite ascriptions would not always be possible, since the relevant typicality contrasts could be either in fact unknown or in principle unknowable. Thus, speakers can agree that there is a contrast in acceptability between (17) and (18), or between (19) and (20),

(17) ? It's wine, but it tastes like wine.
(18)   It's wine, but it doesn't taste like wine.

(19) ? It's a violin, but it produces a violin-like sound.
(20)   It's a violin, but it doesn't produce a violin-like sound.

even when they have neither tasted wine nor heard any violin-like sound (and are, consequently, unable to distinguish between typical and atypical ways wine might taste, or between typical and atypical ways violins might sound). Even more significantly, speakers can agree that there is a contrast in acceptability between (23) and (24), or between (25) and (26),

(23) ? It's a unicorn, but it behaves like a unicorn.
(24)   It's a unicorn, but it doesn't behave like a unicorn.

(25) ? It's a round square, but it looks like one.
(26)   It's a round square, but it doesn't look like one.

even though it is impossible for them to distinguish between typical and atypical unicorn behaviours, or between typical and atypical round square shapes, given that there are no such things as unicorns and round squares at all.

Judging from the two remarks that appear to be meant as replies to this argument, I seriously doubt that Cruse had a clear grasp of its intent. His first remark is that, in the proposed examples, "the expectancies to which *but* is responding are transparently not generated by the relevant lexical items" (Cruse 1992: 400)—a claim that, if true, would hardly be "transparent", and which is anyway false,

as one might confirm by considering the word salads that one would get if one tried to replace, say, "taste" with "sound" in the sentences about wine, or "sound" with "taste" in the sentences about violins. His second, and more revealing, remark is that, in examples of the proposed sort, "the expectancy to which *but* is responding...is simply that a randomly chosen X will more often than not have a lot in common with a typical X" (Cruse 1992: 400). Applied to the examples mentioning unicorns and round squares, this must be taken to mean that, in (23)–(24), 'the expectancy to which *but* is responding is simply that a randomly chosen unicorn will more often than not have a lot in common with a typical unicorn', and that, in (25)–(26), 'the expectancy to which *but* is responding is simply that a randomly chosen round square will more often than not have a lot in common with a typical round square'. The point was, however, that, precisely because unicorns and round squares do not exist, talk of typical and randomly chosen unicorns (and of expected degrees of similarity between them), or of typical and randomly chosen round squares (and of expected degrees of similarity between them), is just literal nonsense, and that one of the best reasons for regarding the argument from *but* as unsuccessful is precisely the fact that its acceptance would force such nonsense upon us. If there are people who can convince themselves that nonsense is sense, they will undoubtedly find no problem here. But those who cannot acquire such convictions must, contrary to Cruse's recommendations, conclude that the problem is so serious as to make any further discussion of the argument from *but* at best otiose.

Chapter 10
Correlative and noncorrelative conjunctions in argument and nonargument positions

1 Introduction

The question as to what the semantic difference is between coordinate structures of the form "X and Y" (hereafter called "noncorrelative conjunctive structures") and coordinate structures of the form "both X and Y" (hereafter called "correlative conjunctive structures") has rarely figured as a topic for discussion, possibly because the definition of conjunction in classical truth-functional logic predisposes one not to expect any semantic difference between the two structures to exist. The present essay makes a preliminary contribution to the investigation of that question, by critically examining an interesting semantic proposal that lies at the core of a recent syntactic analysis of the relation between structures of the two types when they occupy argument positions. I will argue that the semantic proposal in question fails to properly differentiate between the two structures (and that the syntactic claims it was meant to support are consequently undermined). I will also argue, however, that there is clear independent evidence that the two structures *are* semantically distinct, at least when truth-conditionally relevant differences are taken to be—as they generally are taken to be—semantically relevant differences; and that the evidence in question points to the need for a cross-categorial account of that semantic difference, since it is a difference ascertainable both when the structures occupy argument positions and when they occupy nonargument positions.

The analysis to be discussed is due to Progovac (1999) and comprises several theses, most of which cease to be relevant when its central thesis is not granted and need not, therefore, be extensively reviewed in what follows. The central thesis is an answer to the question as to how the semantic difference between structures of the form "X and Y" and corresponding structures of the form "both X and Y", when both structures occupy argument positions, is to be characterised. Briefly, Progovac's answer to that question is as follows: In a correlative conjunctive structure, the conjuncts are necessarily interpreted as referring to two distinct occupants of a given θ-role (e.g., to two distinct agents, two distinct themes, two distinct goals, etc.) As a consequence, the sentence where the conjunctive structure occurs is necessarily interpreted as referring to two distinct events or

states (given that a single event or state may not comprise, according to Progovac, more than one occupant of any given θ-role). In a noncorrelative conjunctive structure, on the other hand, the conjuncts may be interpreted as referring either to two distinct occupants of a given θ-role or to a single, "collective", occupant of a given θ-role. As a consequence, the sentence where the conjunctive structure occurs may be interpreted as referring either to two distinct events or states or to a single event or state. Thus, in (1) and (2),

(1)   I visited [Maria and Peter]
(2)   I gave a rose to [Maria and Peter]

the coordinate structures may be interpreted, according to Progovac, either as referring to two distinct occupants of a given θ-role—two distinct themes in the case of (1) and two distinct goals in the case of (2)—or as referring to a single, "collective", occupant of a given θ-role—a single theme in the case of (1) and a single goal in the case of (2). As a consequence, each sentence may be interpreted either as referring to two different events—two different visiting events in the case of (1) and two different rose-giving events in the case of (2)—or to a single event— a single visiting event in the case of (1) and a single rose-giving event in the case of (2). On the other hand, in (3) and (4),

(3)   I visited [both Maria and Peter]
(4)   I gave a rose to [both Maria and Peter]

the coordinate structures are necessarily interpreted, according to Progovac, as referring to two distinct occupants of a given θ-role—two distinct themes in the case of (3) and two distinct goals in the case of (4). And because multiplicity of a given θ-role's occupants implies multiplicity of events, each sentence is necessarily interpreted as referring to two distinct events rather than to a single event— two distinct visiting events, rather than a single visiting event, in the case of (3); and two distinct rose-giving events, rather than a single rose-giving event, in the case of (4).

Assuming that this semantic characterisation is correct, Progovac then sets out to exploit it syntactically, by arguing that it supports the hypothesis that "X and Y" and "both X and Y" are derivationally related and that their derivational relation is constrained by an economy principle of the sort favoured by the Minimalist Program—a hypothesis that would, of course, be significant if the semantic characterisation on which it is explicitly based was secure, but that cannot be

accepted if that semantic characterisation is, as I am now going to argue, seriously problematic.

2 Two problems

Progovac's view that, when "X and Y" and "both X and Y" occupy argument positions, they are, by virtue of their structural differences, always semantically distinct is spelled out as the thesis that, whereas a correlative structure is necessarily interpreted as referring to two distinct occupants of a given θ-role, and so to participants of two distinct events, a noncorrelative structure may be interpreted either as referring to two distinct occupants of a given θ-role, and so to participants of two distinct events, or to a single occupant of a given θ-role, and so to a single event. (The relevant distinction will henceforth be referred to as the distinction between "single-event" and "multiple-event" interpretations.) To show that this characterisation of the difference between "X and Y" and "both X and Y" is incorrect, one could, accordingly, do the one or the other of two things (among others). First, one could show that there are pairs of sentences both of whose members can *only* receive the multiple-event interpretation, even though their unique difference is that the one contains a correlative structure and the other the corresponding noncorrelative structure. Second, one could show that there are pairs of sentences both of whose members can *only* receive the single-event interpretation, even though their unique difference is that the one contains a correlative structure and the other the corresponding noncorrelative structure. Both of these things, however, can be established fairly straightforwardly.

Consider first the following sentences, each of which is grammatical and none of which is contradictory:

(5)   a.   Last week, John travelled from Paris to Berlin and from Berlin to Paris.

       b.   Last week, John travelled both from Paris to Berlin and from Berlin to Paris.

(6)   a.   Last night, John felt the rise and the fall of the room's temperature.

       b.   Last night, John felt both the rise and the fall of the room's temperature.

It may not be clear what the semantic difference between the members of each pair is, but it is clear that it cannot be the difference predicted by Progovac's analysis. No matter what theory of events one assumes, a person's travelling from Paris to Berlin cannot be the very same event as that person's travelling from Berlin to Paris, and so the only interpretation that either (5a) or (5b) can receive is the multiple-event interpretation. Notice, in this connection, that both sentences, and not just (5b), become contradictory when supplied with an adverb that would be ideal for releasing a coherent single-event interpretation, if that interpretation were indeed available (here and elsewhere in this chapter, contradictoriness is marked by "#"):

(7)   a.   #   Last week, John simultaneously travelled from Paris to Berlin and from Berlin to Paris.
      b.   #   Last week, John simultaneously travelled both from Paris to Berlin and from Berlin to Paris.

Similarly, no matter what theory of events one assumes, a person's feeling a room's temperature rise cannot be the very same event as that person's feeling the same room's temperature fall, and so the only interpretation that both (6a) and (6b) can receive is the multiple-event interpretation. Notice, again, that both sentences, and not just (6b), become contradictory when supplied with an adverb that would be ideal for releasing a single-event interpretation, if that interpretation were in fact possible:

(8)   a.   #   Last night, John felt simultaneously the rise and the fall of the room's temperature.
      b.   #   Last night, John felt simultaneously both the rise and the fall of the room's temperature.

Progovac's account, however, clearly predicts that, although the multiple-event interpretation is the unique interpretation available to the (b)-members of these pairs, both the multiple-event *and* the single-event interpretations are always freely available to the (a)-members of the same pairs. In fact, any alternative predictions would not only remove any basis for sustaining her claim that correlative and noncorrelative conjunctive structures are semantically distinct just by virtue of their structural differences, but would also openly contradict the economy-driven account that she proposes for the syntactic relationship between them. On that account, the reason why correlative conjunctive structures exist alongside corresponding non-correlative conjunctive structures is that the greater syntactic

cost involved in the derivation of the former, as compared to the derivation of the latter, is compensated by the fact that the former are more selective in the number of interpretations they allow than the latter.[1] And since the greater semantic selectivity of a correlative as opposed to a corresponding noncorrelative conjunctive structure consists, according to Progovac, in the fact that the correlative one is only compatible with multiple-event interpretations whereas the noncorrelative one always allows either single-event or multiple-event interpretations, it follows that a sentence containing a noncorrelative conjunctive structure could not possibly exist alongside a corresponding sentence containing a correlative conjunctive structure *unless* the former sentence was semantically *less* selective than the latter—that is, unless the noncorrelative one admitted either single-event or multiple-event interpretations whereas the correlative one admitted only multiple-event interpretations. However, the (a)-members of these sentence pairs, which contain noncorrelative structures, admit only multiple-event interpretations, just like the corresponding (b)-members, which contain correlative structures. They therefore falsify the semantic characterisation of the difference between correlative and noncorrelative structures and, along with it, the economy-driven characterisation of their syntactic relationship.[2]

The inadequacy of Progovac's account, however, derives not only from the fact that there exist noncorrelative conjunctive structures that happen to induce exclusively multiple-event interpretations, but also from the fact that there exist correlative conjunctive structures that happen to induce exclusively single-event interpretations. Consider the following sentences, each of which is grammatical and none of which is contradictory:

---

1 Correlative structures are syntactically more costly than noncorrelative ones, according to Progovac, because they involve merging of pronounced categories whereas the latter involve merging of silent categories.

2 Notice that it would not do to suggest that, appearances notwithstanding, (5a) and (6a) really are ambiguous, as Progovac predicts, between the single-event and the multiple-event interpretations, but that their ambiguity is masked by "pragmatic factors". The problem with this suggestion is that, if it were correct, (7a) and (8a) would not be contradictory but merely pragmatically odd. Clearly, however, (7a) and (8a) are contradictory rather than merely pragmatically odd: it is not unusual or inconvenient but *impossible* for someone to *simultaneously* travel from Paris to Berlin and from Berlin to Paris; and it is not unusual or inconvenient but *impossible* for someone to *simultaneously* feel a certain room's temperature rise and fall. Therefore, (5a) and (6a) are confined to the multiple-event interpretation just as firmly as (5b) and (6b) are, and the suggested "pragmatic" way out of the difficulty leaves the difficulty exactly where it was.

(9) a. Since the person you most admired and the person you most feared was the very same person, by killing that person, you killed both the person you most admired and the person you most feared.

b. Since the person you most admired and the person you most feared was the very same person, by killing that person, you killed the person you most admired and the person you most feared.

(10) a. Since the fruit you most like and the fruit that is worst for your health is the same fruit, by eating that fruit, you ate both the fruit you most like and the fruit that is worst for your health.

b. Since the fruit you most like and the fruit that is worst for your health is the same fruit, by eating that fruit, you ate the fruit you most like and the fruit that is worst for your health.

Whatever the semantic difference between the members of each pair might be, it is certainly not the difference predicted by Progovac. There can be no credible theory of events from which it would follow that, if you have killed a person that happens to be both admired and feared by you, then you have committed *two* killings, or from which it would follow that, if you have eaten something that is both likeable to you and detrimental to your health, then you have eaten *two* different things. (Indeed, Davidson's theory of events, to which Progovac declares her allegiance at the very beginning of her paper, is specifically designed to block absurd implications of this sort, among others.[3]) If this so, however, the only interpretation that both members of each pair can receive is the single-event interpretation—that is, the interpretation according to which a *single* killing event is reported in both (9a) and (9b) and a *single* eating event is reported in both (10a) and (10b). But this contradicts Progovac's semantic thesis that sentences containing correlative conjunctive structures are necessarily interpreted as referring to multiple events, whereas corresponding sentences containing noncorrelative conjunctive structures may be interpreted as referring either to a single event or to multiple events. And given that the economy-driven account proposed by Progovac for the syntactic relationship between correlative and non-correlative conjunctive structures crucially depends on the assumption that this semantic

---

3 For the basic features of Davidson's theory, see the papers collected in Davidson (1980). For discussions and developments, see the papers collected in Lepore and McLaughlin (1985) and in Vermazen and Hintikka (1985).

difference always exists, it follows that neither the semantic nor the syntactic proposal that she offers can be accepted.

To summarise, Progovac's account of the relation between correlative and noncorrelative conjunctive structures in argument positions is not successful, first because it predicts single-event interpretations where none is possible, and secondly because it predicts multiple-event interpretations where none is possible. Assuming, therefore, that Progovac would not wish to claim that her analysis both incorporates an analysis of events and is entirely free from any independently justified constraints on event individuation (a hardly attractive position, even apart from the circularity it would involve), her proposal cannot be taken, in its present form, to have satisfactorily attained its intended goal.

3 Two questions and a further problem

Rather than expediting the problem of the proper characterisation of the relation between correlative and noncorrelative structures to the repository of problems awaiting solution, it might now be interesting to raise two questions. First, is there an *alternative* basis for claiming that the two types of structure are semantically distinct in argument positions, given that their difference in such positions cannot always be traced, as Progovac was supposing, to the distinction between single and multiple events? Second, is there a basis for supposing, contrary to what Progovac must be supposed to have been assuming, that the semantic difference between correlative and non-correlative structures is independent of whether these structures occupy argument or nonargument positions, and that, therefore, a uniform account of it should be envisaged as a theoretical possibility? I wish to suggest that affirmative, and closely related, answers to both questions appear to be available. In the course of developing these answers, a further problem facing Progovac's original analysis will be seen to emerge.

With respect to the first question, it is clear that the semantic distinctness of correlative and noncorrelative structures in argument positions can be established without invoking the opposition between single and multiple events, by simply noticing that there are types of sentential contexts (specifically, contexts characterised by the presence of certain temporal or deontic operators) within which intersubstitution of noncorrelative and correlative structures produces a difference in truth value. Thus, the fact that (11) and (12) are not semantically equivalent

(11)   John gave Mary [candies and nuts]

(12)   John gave Mary [both candies and nuts]

can be reliably inferred, given standard assumptions about semantic compositionality, from the fact that sentences like (13) or (15) are not contradictory, whereas sentences like (14) or (16) are certainly contradictory.[4]

(13)   John gave Mary candies and nuts on several occasions during the year, but it is only on Christmas Eve that he gave her both candies and nuts.
(14) # John gave Mary candies and nuts on several occasions during the year, but it is only on Christmas Eve that he gave her candies and nuts.
(15)   John may give Mary candies and nuts if he likes, but he must never give her both candies and nuts.
(16) # John may give Mary candies and nuts if he likes, but he must never give her candies and nuts.

Assuming, as it is generally assumed, that a truth-conditionally relevant difference is a semantic difference, such contrasts amply justify the hypothesis that correlative and noncorrelative structures in argument positions are semantically distinct, without making it necessary to invoke the opposition between single-event and multiple-event interpretations, even if the application of that opposition were not problematic, as we independently know it is.

Indeed, the sorts of operators that trigger the truth-conditional differences between correlative and noncorrelative structures not only make it unnecessary to invoke an opposition between single-event and multiple-event interpretations, but are also instrumental in showing that the terms of that opposition are often impossible to apply to the relevant data, thus providing a further argument against Progovac's analysis. Consider (17)–(20), each of which contains a correlative conjunctive structure and each of which would therefore be predicted by Progovac to refer necessarily to multiple events or states rather than to a single event or state.

(17)   I gave a lesson to both Mary and Carol.
(18)   I was attacked by both Mary and Carol.
(19)   I danced with both Mary and Carol.

---

4 All occurrences of "and" in examples (13)–(16) are assumed not to be emphatically stressed, thus ensuring that the contradictoriness and non-contradictoriness judgments are responsive specifically to semantic properties of the conjunction and not to additional semantic properties of emphatic stress.

(20)   I was married to both Mary and Carol.

Including these sentences in the scope of appropriate temporal or deontic opera-
tors, however, suggests that the alleged multiple-event or multiple-state interpre-
tations not only are not necessary but may on occasion be impossible. For exam-
ple, each of the following sentences can perfectly well be true:

(21)   I can give a lesson to Mary, I can give a lesson to Carol, but I cannot give
        a lesson to both Mary and Carol.
(22)   I was once attacked by Mary, I was once attacked by Carol, but I was
        never attacked by both Mary and Carol.
(23)   I may dance with Mary, I may dance with Carol, but I must never dance
        with both Mary and Carol.
(24)   I was once married to Mary, I was once married to Carol, but I was never
        married to both Mary and Carol.

Clearly, none of these sentences could be true unless the event or state that each
one of them describes as *not* having occurred or *not* being allowed was a *single*
event or state. After all, the point of each sentence is to contrast the non-occur-
rence or non-permissibility of that single event or state with the occurrence or
permissibility of corresponding *separate* events or states. For example, (23)
claims that although the speaker is not entitled to a single dancing event with
two partners, he *is* entitled to two separate dancing events with single partners;
and (24) claims that although the speaker was not involved in a single bigamous
state, he *was* involved in two separate monogamous states. However, the partic-
ipants of the *single* event or state to which each of the sentences refers—in order
to contrast it with corresponding separate events or states—are described in the
sentences by means of a *correlative* conjunctive structure, which, on Progovac's
analysis, necessarily refers to participants of *separate* events or states rather than
to participants of a single event or state. Given, then, that—contrary to what
Progovac's account would predict—none of the sentences is contradictory, an ad-
equate analysis of the correlative conjunctive structures they contain could not
possibly require that these structures necessarily refer to participants of multiple
events or states rather than to participants of single events or states. And if cor-
relative structures can refer either to single or to multiple events or states, they
can hardly be distinguished, on the basis proposed by Progovac, from noncorrel-
ative ones, which, in her view, can do either of these things just as well. The two
types of structure, however, certainly remain truth-conditionally distinct for the
reasons already indicated, and it is therefore their truth-conditional difference,

rather than the suggested event-theoretic considerations, that ensure their semantic distinguishability.[5]

Regarding the second of the questions raised above, it is clear that the kinds of sentential contexts that can be used for establishing that correlative and noncorrelative structures in argument positions are truth-conditionally distinct can *also* be used for establishing that correlative and noncorrelative structures in *nonargument* positions are truth-conditionally distinct, too—which suggests that, unless specific arguments to the contrary were forthcoming, the search for an account that would provide a unified treatment of the semantic opposition between the structures *irrespective* of whether or not they occupy argument positions should be given methodological priority. For example, the question of semantic distinctness can be asked about (25) and (26), where the conjunctive structures occupy nonargument positions, just as properly as it has been asked about (11) and (12), where the conjunctive structures occupy argument positions.

(25)   John [danced and played] with Mary.
(26)   John [both danced and played] with Mary.

And just as an affirmative answer to that question with respect to (11) and (12) has been possible by noticing the discriminating effects of temporal and deontic operators, an affirmative answer to the same question with respect to (25) and (26) is possible by noticing the discriminating effects of the same temporal and deontic operators: The fact that (25) and (26) are semantically distinct can be reliably inferred, given standard assumptions about semantic compositionality, from the

---

5 It is worth noting that the argument developed here is also applicable to the very examples that Progovac analyses. Thus, her claim that, in "I gave a rose to both Maria and Peter" (=(2) above), the speaker necessarily refers to two separate rose-giving events makes it impossible to explain why the following sentence can actually be true:

> I once gave a rose to Maria, I once gave a rose to Peter, but—in order to prevent quarrels between them—I never gave a rose to both Maria and Peter.

Obviously, this sentence could not be true unless the rose-giving event the speaker claims to have avoided was a *single* rose-giving event involving both Maria and Peter rather than two separate rose-giving events having Maria and Peter as their respective goals (these separate events are, in fact, explicitly claimed to actually have occurred and not to be among the events the speaker was at all concerned to avoid). However, the participants of this *single* rose-giving event are referred to in the sentence by means of a *correlative* conjunctive structure. Therefore, the fact that the sentence can be true contradicts Progovac's generalisation that correlative structures can *only* refer to participants of multiple events.

fact that sentences like (27) and (29) are not contradictory, whereas sentences like (28) and (30) are certainly contradictory.[6]

(27)   John danced and played with Mary on several occasions during the year, but it is only on Christmas Eve that he both danced and played with her.
(28) # John danced and played with Mary on several occasions during the year, but it is only on Christmas Eve that he danced and played with her.
(29)   John may dance and play with Mary if he wants to, but he must never both dance and play with her.
(30) # John may dance and play with Mary if he wants to, but he must never dance and play with her.

If, however, it is the same types of linguistic triggers that bring to the fore the semantic difference between correlative and noncorrelative structures *whether or not* these structures occupy argument positions, it would certainly be premature to assume that the treatment of this difference as it manifests itself in argument positions should or could be attempted quite independently of any investigation of its possible manifestations in nonargument positions. The methodologically advisable option would be, rather, to try to capture that difference in terms independent of the distinction between argument and nonargument positions, leaving the possibility of separate treatments as a last resort, to be used only when a unified account becomes demonstrably unattainable.

4 Conclusion

Clearly, much further work on the various aspects of the relation between correlative and noncorrelative structures remains to be done. Nevertheless, some conclusions seem already securely derivable on the basis of the preceding discussion.

As far as the semantic relationship between the two types of structure is concerned, the conclusion is twofold. First, there is clear evidence suggesting that correlative and noncorrelative structures are truth-conditionally distinct, and therefore—assuming generally held assumptions about the correlation between

---

6 All occurrences of "and" in examples (27)–(30) are assumed not to be emphatically stressed, thus ensuring that the contradictoriness and non-contradictoriness judgments are responsive specifically to semantic properties of the conjunction and not to additional semantic properties of emphatic stress.

truth-conditional relevance and semantic relevance—semantically distinct. Given that the properties of the contexts that trigger that difference are the same whether or not the structures in question occupy argument positions, it is highly plausible to assume, until strong evidence to the contrary is accumulated, that their semantic difference is *independent* of whether or not they occupy such positions and should be given an account that is cross-categorially applicable. Second, even if it were discovered that, for some presently unobvious reason, the semantic difference between correlative and noncorrelative structures in argument positions should be accounted quite independently of their semantic difference in nonargument positions, no viable account of their semantic difference in argument positions can be couched in terms of the opposition between singular and non-singular events, because there is, in fact, no reliable correlation between the presence of a correlative or noncorrelative structure in a sentence and the singularity or non-singularity of the event reported by that sentence.

Concerning the syntactic relation between correlative and noncorrelative conjunctive structures, the main conclusion deriving from the present discussion is that, whatever that syntactic relation is, it is not subject to an economy constraint of the kind advocated by Progovac. Whether or not this fact militates against the aspect of the Minimalist Program that Progovac was engaged in developing depends on whether or not that program should be committed to the search of economy constraints of that sort. Along with Progovac, I would be inclined to believe that it should—though, given the unclarity of the theoretical situation in that area, I prefer to leave it to the readers to position themselves on this issue as best as each of them can.

Part II **Matters of meaning and force**

Chapter 11
Yes–no questions and the myth of content invariance

1 Two kinds of force-content distinction

No theory of sentence meaning would be adequate if it failed to entail that a non-declarative sentence like *Is water odourless?* and a declarative sentence like *Water is odourless*, though both meaningful, do not have the same meaning; and only theories of meaning that, like Searle's, aim to systematically relate differences in sentence meaning to differences in illocutionary act potential would have any chance of engendering such entailments. Still, not all ways of relating sentence meanings to illocutionary acts are adequate, and in this essay I want to argue that a fundamental assumption that Searle uses in analysing sentence meaning in terms of illocutionary acts is mistaken. The assumption (which is very widely shared among those who, along with Searle, duly acknowledge that no account of sentence meaning can dispense with an account of sentence mood) has to do with the particular way in which Searle interprets the distinction between the *force* and the *content* of illocutionary acts and applies it to the analysis of sentence meaning.

There is an innocuous way of interpreting the force-content distinction against which there can be no objection, and which I would be perfectly happy to accept. On that innocuous interpretation, saying that illocutionary acts (and the sentence meanings that one aims to characterise in their terms) should be thought of as consisting in the attachment of a *force* to a *content* simply means that, in order to fully specify an illocutionary act in a semantically relevant way, it is not enough to merely specify the *kind* of illocutionary act it is (whether, for example, it is a yes–no question, or an assertion, or a command, etc.), but it is also necessary to specify what it is about its meaning that distinguishes it from all other illocutionary acts of the *same* kind (for example, what it is about the meaning of a yes–no question that makes it different from another yes–no question, what it is about the meaning of an assertion that makes it different from another assertion, what it is about the meaning of a command that make it different from another command, and so on). That both of these kinds of specification are required is obvious, but merely registering that requirement in the way the innocuous interpretation does is not tantamount to adopting any *particular* way

in which one might propose to fulfil it. As far as the innocuous interpretation is concerned, the force of an illocutionary act is *whatever* turns out to properly distinguish that act, in semantically relevant terms, from all illocutionary acts of *other* kinds, and the content of an illocutionary act is *whatever* turns out to properly distinguish that act, in semantically relevant terms, from all illocutionary acts of the *same* kind. But accepting the force-content distinction in that innocuous sense involves no strong commitments as to what sorts of things forces or contents will turn out to be, and is, in particular, perfectly consistent with the possibility that there may not be a single *kind* of content uniformly attributable *across* illocutionary act kinds—it is perfectly consistent, for example, with the possibility that what makes a yes–no question semantically distinct from another yes–no question is a different *sort* of thing from what makes an assertion semantically distinct from another assertion, which in turn is a different *sort* of thing from what makes a command semantically distinct from another command, and so on.

In contrast to the innocuous interpretation, however, the standard interpretation of the force-content distinction, of which Searle is a paradigmatic representative, involves quite strong commitments both as to what forces are and as to what contents are. And it is an implication of its commitments in the latter area that there does exist a single *kind* of content that is uniformly attributable across illocutionary act types in all theoretically fundamental cases (i.e. in all cases in which the content in question is representable by means of a complete sentence), and that ensures, among other things, that the *sort* of thing that makes a yes–no question semantically distinct from another yes–no question is the same as the *sort* of thing that makes an assertion semantically distinct from another assertion, which in turn is the same as the *sort* of thing that makes a command semantically distinct from a another command, and so on. In Searle's case (which, in this respect too, is fully representative of standard assumptions) the unique sort of thing that plays the role of illocutionary act *content* in all theoretically central cases is a *proposition*, and it is the differences between the *propositions* to which they associate the same force that accounts for differences in meaning between illocutionary acts of the same kind. Furthermore, just as there are differences in meaning between illocutionary acts that are due exclusively to the fact that different *propositions* get attached to the same force, there are also, in Searle's view, differences in meaning between illocutionary acts that are due exclusively to the fact that different forces get attached to *the same proposition*. And it is differences of that last kind that, according to Searle and many others, account for the distinctive semantic contribution of sentence mood: semantically distinct sentences that differ only in mood are sentences whose semantic non-identity is uniquely

attributable to the fact that each of them associates a different kind of force to the very same *proposition*.

I will argue that these views cannot be maintained if some familiar and fundamental facts regarding the interpretation of yes–no questions are to be respected. And since a theory of sentence mood—and so, of sentence meaning—that cannot deliver a satisfactory analysis of as simple and basic a sentence type as a yes–no question cannot be accepted, I will conclude that no theory of sentence meaning will be adequate if it incorporates the standard interpretation of the force-content distinction, and in particular its assumption that the contents to which illocutionary forces are attached are propositional. Along the way, I will be gesturing towards the sort of analysis of yes–no questions that I would regard as appropriate; my primary purpose, however, is not to fully work out such an analysis, but rather to argue that certain features of yes–no questions that *any* analysis should accommodate are not consistent with certain basic assumptions that many theories of meaning routinely endorse.

2 Yes-no questions and the analysis of illocutionary acts

According to Searle, the literal meaning of a sentence derives from (in the sense, presumably, of being a logical construction out of) the illocutionary acts that linguistic conventions allow its speakers to perform in uttering it, and a typical illocutionary act, noted *F(p)*, consists of an illocutionary force *F* attached to a proposition *p*, which represents the act's content: "throughout the analysis of speech acts", Searle explains, he is observing a distinction between "what might be called *content* and *function*", where "the content is the proposition" and "the function is the illocutionary force with which the proposition is presented" (Searle 1969: 125).

The principal virtue of the distinction between illocutionary force and propositional content as the two semantically relevant components of an illocutionary act is, Searle suggests, that it enables us to appreciate that differences in meaning between illocutionary acts can result either from keeping an illocutionary force constant and varying the propositions to which that force is attached or from keeping a propositional content constant and varying the illocutionary forces attached to it (and it is, as already mentioned, the latter possibility that is the key to Searle's—and to many others'—account of sentence mood). Thus, the sentences *Water is odourless* and *Blood is colourless* differ in meaning because the

illocutionary acts that linguistic conventions allow their speakers to perform are different by virtue of attaching the same force (the force of assertion) to two *different* propositions (the proposition that water is odourless and the proposition that blood is colourless, respectively). On the other hand, the sentences *Water is odourless* and *Is water odourless?* differ in meaning because the illocutionary acts that linguistic conventions allow their speakers to perform are different by virtue of attaching two *different* forces (the force of assertion and the force of question, respectively) to the *same* proposition (the proposition that water is odourless); and similarly, the sentences *Blood is colourless* and *Is blood colourless?* differ in meaning because the illocutionary acts that linguistic conventions allow their speakers to perform are different by virtue of attaching two different forces (the forces of question and assertion, respectively) to the same proposition (the proposition that blood is colourless).

The relation, illustrated by the last two pairs of examples, between a yes–no question and its grammatically corresponding assertion, has always been presented by Searle as an exceptionally clear case of a relation between sentences that are conventionally dedicated to the performance of acts with *different* illocutionary forces but the *same* propositional content, and this fact fully justifies attributing to him the following thesis, which will be central to the present discussion, and which I will call the Content Invariance Thesis:

> *Content Invariance Thesis (CIT)*
>
> Under uniform interpretations of corresponding sub-sentential constituents, a yes–no question and its grammatically corresponding assertion express the same proposition.

Thus, in discussing the examples *Sam smokes habitually* and *Does Sam smoke habitually?*, which are the very first examples by means of which the distinction between illocutionary force and propositional content is introduced in *Speech Acts*, Searle explicitly claims that, in utterances of them where the same individual, Sam, is being referred to, "the same proposition is expressed" (Searle 1969: 29)—the proposition, namely, that Sam smokes habitually. In his discussion of predication later in the same book (Searle 1969: 124), he explicitly assumes that "the man who asserts that Socrates is wise [and] the man who asks whether he is wise" express the same proposition—the proposition, namely, that Socrates is wise. In his discussion of Russell's theory of descriptions toward the end of the same book (Searle 1969: 150–62), he maintains that no theory of descriptions would be adequate if it could not acknowledge the fact that *the same proposition*

is expressed by the assertion *The King of France is bald* and by the yes–no question *Is the King of France bald?* In discussing the problem of opacity in *Intentionality* (Searle 1983: 180–96), he remarks that, assuming that their subject terms refer to the same individual, the assertion *Mr Howard is an honest man* and the yes–no question *Is Mr Howard an honest man?* have the same propositional content—the content, namely, that Mr Howard is an honest man—and emphasises that exactly the same content is expressed by the complement clauses of their respective indirect speech reports (that is, by the complement clauses of such sentences as *X asserted that Mr Howard is an honest man* and *X asked whether Mr Howard is an honest man*), notwithstanding the fact that different complementisers—"that" and "whether" respectively—would be grammatically required to introduce the two complement clauses in question. And in a more recent book summarising some of his main philosophical contributions (Searle 1998), he begins his discussion of language by drawing attention to the significance of the distinction between the force and the content of illocutionary acts, offering as an "obvious example" of cases illustrating that distinction the assertion *You will leave the room* and the yes–no question *Will you leave the room?*, which, he remarks, are clearly different in one respect (namely, in the force that each one carries) but, equally clearly, identical in another respect, namely, in the proposition that each one expresses, which is "the proposition that you will leave the room" (Searle 1998: 138; the same examples, reported in indirect speech, are used to make the same point in Searle 2004: 166). There can be no doubt, then, that, according to Searle, *the same proposition* is expressed by a yes–no question and by its grammatically corresponding assertion (as well as by the complement clauses of correct indirect speech reports of these), under uniform interpretations of their corresponding subsentential constituents. And in holding this Searle is, of course, far from alone: terminological differences aside, what the Content Invariance Thesis proposes is accepted as a matter of course by several other prominent philosophers—perhaps not surprisingly, since the original formulation of the thesis occurs, in the context of a general discussion of the force-content distinction, in a late essay by Frege (1918) that is widely regarded as a founding document of modern philosophy of language.

I am going to argue that no acceptable account of either propositions or illocutionary acts is consistent with adherence to the Content Invariance Thesis, and that that thesis should consequently be rejected. (Rejecting CIT need not, of course, involve commitment to the thesis that yes–no questions have *different* propositional contents from their grammatically corresponding assertions: it may rather involve commitment to the thesis that, unlike their grammatically corresponding assertions, yes–no questions have no *propositional* contents at all.)

The argument relies on two theses that neither Searle nor anyone else would presumably have the desire to controvert, and consists in showing that, when taken in conjunction with these uncontroversial theses, CIT has a range of implications that turn out to be obviously false. The two uncontroversial theses, which I will respectively call the Propositional Distinctness Thesis and the Illocutionary Distinctness Thesis, are the following:

> *Propositional Distinctness Thesis (PDT)*
>
> Two propositions are not identical if the one is true and the other false with respect to the same situation.

> *Illocutionary Distinctness Thesis (IDT)*
>
> Two illocutionary acts are not identical if they result from attaching the same force to two propositions that are not identical.

In the sense relevant to the present discussion, propositions are entities that, by definition, have truth conditions (''to know the meaning of a proposition'', Searle acknowledges [1969: 125], ''is to know under what conditions it is true or false''); and although not everyone would accept that it is sufficient for two propositions to have the same truth conditions in order for them to be identical, absolutely no one would deny that it *is* sufficient for two propositions *not* to have the same truth conditions in order for them *not* to be identical. The PDT thesis simply states that universally accepted condition, and should therefore be embraced by anyone who cares to speak about propositions at all: if, with respect to *the same* situation, one proposition is true and another false, then they are certainly not *the same* proposition. As for the IDT thesis, it could hardly be denied by anyone who contemplates importing talk of propositions into the analysis of illocutionary acts: if attaching its force to a particular proposition is rightly taken to be essential to an illocutionary act's identity, then attachments of nonidentical propositions to a given illocutionary force should be expected to make a crucial difference to the identity of the resulting illocutionary acts, and the IDT thesis simply spells out the expected crucial difference—namely, that if two illocutionary acts consist in attaching the same force to two propositions that are *not* identical, then the illocutionary acts *themselves* will not be identical.

We are now going to see that, in conjunction with CIT, the two uncontroversial theses just stated have implications that, though not false in the case of assertions, are clearly false in the case of yes–no questions. The few examples to be discussed will be easy to generalise in obvious ways. Assuming uniform interpretations of subsentential constituents, the assertion in (1) and the yes–no question

in (2) express, according to CIT, the *same* proposition—namely, the proposition that Sam is autistic:

(1)   Sam is autistic.
(2)   Is Sam autistic?

And, assuming again uniform interpretations of subsentential constituents, the assertion in (3) and the yes–no question in (4) express, according to CIT, the *same* proposition—namely, the proposition that Sam is not autistic:

(3)   Sam isn't autistic.
(4)   Isn't Sam autistic?

Now, the proposition that Sam is autistic and the proposition that Sam is not autistic are certainly not identical propositions, for the kind of reason that PDT states—namely, because their truth conditions are not identical. But then, given IDT, one should expect these nonidentical propositions to give rise to nonidentical illocutionary acts, when the same force attaches to them. And that turns out *not* to be invariably the case: utterances in which the two propositions are associated with the force of *assertion* do indeed instantiate *different assertions*, rather than different *ways* of making the *same* assertion: clearly, (1) and (3) are not different ways of making the *same* assertion, they are simply different, indeed incompatible, assertions:

(1)   Sam is autistic.
(3)   Sam isn't autistic.

However, utterances in which the two propositions are putatively associated with the force of *question* do *not* instantiate different questions, but merely different *ways* of posing *the same question*: clearly, (2) and (4) do not pose different questions each, but merely constitute two different *ways* of posing *the same question*:

(2)   Is Sam autistic?
(4)   Isn't Sam autistic?

That (2) and (4) instantiate the same question even though (1) and (3) instantiate different assertions is evident from the fact that, although the affirmations and

denials of identity that (5) and (7) contain are obviously true, there is no interpretation in which the affirmations and denials of identity that (6) and (8) contain are true:

(5)   I already asked you whether Sam is autistic, and I am now asking you the same thing again: isn't he autistic?

(6)   #   I already told you that Sam is autistic, and I am now telling you the same thing again: he isn't autistic.

(7)   I did tell you before that Sam is autistic, but what I am now telling you is different: he isn't autistic.

(8)   #   I did ask you before whether Sam is autistic, but what I am now asking you is different: isn't he autistic?

These contrasts are impossible to explain unless one assumes that, whereas *Sam is autistic* and *Sam isn't autistic* instantiate different assertions (and not different *ways* of making the same assertion), *Is Sam autistic?* and *Isn't Sam autistic?* do *not* instantiate different questions (but merely different ways of posing the *same* question). Furthermore, if *Is Sam autistic?* and *Isn't Sam autistic?* were instantiating different questions, it should be possible for one of those questions to be *settled*—that is, to have its correct answer selected among relevant alternatives—*without* the other being thereby settled. But that is simply impossible, as the following contradictory statements illustrate:

(9)   The question whether Sam is autistic has now been settled, but it is not yet settled whether he isn't autistic.

(10)   The question whether Sam isn't autistic has now been settled, but it is not yet settled whether he is autistic.

And, of course, if (2) and (4) were instantiating different questions, it should be possible for someone to come to know, or to need to know, the answer to the one *without* thereby coming to know, or having the need to know, the answer to the other. But that, again, is simply impossible, as the following incoherent pronouncements indicate:

(11)   I know very well whether Sam is autistic, but what I still don't know is this: is he not autistic?

(12)   I have no need to know whether Sam is not autistic, but what I do need to know is this: is he autistic?

In short, and contrary to what one should expect given CIT and the undisputed theses PDT and IDT, it is just as clear that (2) and (4) instantiate the *same* question as it is clear that (1) and (3) do *not* instantiate the same assertion. And this, of course, is hardly an isolated fact. Any pair of *declarative* sentences of which the one combines a singular term with a given predicate and the other the same singular term with the negation of that predicate will be a pair of sentences whose members are conventionally dedicated to instantiating *different*, indeed incompatible, assertions. On the other hand, any pair of *interrogative* sentences of which the one combines a singular term with a given predicate and the other the same singular term with the negation of that predicate will be a pair of sentences whose members are conventionally dedicated to instantiating the *same* question. And, just as one would expect, the same kind of contrast can be observed when the relevant declarative and interrogative sentences are of greater logical complexity. For example, the proposition that there is nothing divisible by zero and the proposition that there is something divisible by zero are certainly not identical propositions, for the kind of reason stated in PDT—namely, because their truth conditions are different. In conjunction with IDT, then, the Content Invariance Thesis would entail not only that the result of attaching the force of *assertion* to these nonidentical propositions would be two different assertions, but also that the result of attaching the force of *question* to these nonidentical propositions would be two different questions. But that turns out *not* to be the case. When the force that is being attached is the force of assertion, the resulting illocutionary acts are indeed different—interpreted with respect to the same quantificational domain, the declarative sentences in (13) and (14) instantiate different, indeed incompatible, assertions, and not different *ways* of making the *same* assertion:

(13)   There is nothing divisible by zero.
(14)   There is something divisible by zero.

When, however, the force that is being attached is the force of question, the resulting illocutionary acts are identical—interpreted with respect to the same quantificational domain, the interrogative sentences in (15) and (16) do *not* instantiate different questions, but merely different *ways* of posing the *same* question:

(15)   Is there nothing divisible by zero?
(16)   Is there something divisible by zero?

And that, as before, is easy confirmable. Sentences such as those in (17) and in (18) are obviously contradictory:

(17)   The question whether there is something divisible by zero is not yet settled, but it is already settled whether there is nothing divisible by zero.

(18)   The question whether there is nothing divisible by zero is not yet settled, but it is already settled whether there is something divisible by zero.

And although the affirmations and denials of identity that (19) and (21) contain are obviously true, there exist no interpretations of (20) and (22) in which the affirmations and denials of identity that *they* contain are true:

(19)   I already asked you whether there is nothing divisible by zero, and I am now asking you the same thing again: is there something divisible by zero?

(20) # I already told you that there is nothing divisible by zero, and I am now telling you the same thing again: there is something divisible by zero.

(21)   I did tell you that there is nothing divisible by zero, but what I am now telling you is different: there is something divisible by zero.

(22) # I did ask you whether there is nothing divisible by zero, but what I am now asking you is different: is there something divisible by zero?

It must simply be accepted, therefore, that, given two representations that, construed as propositions, would be inconsistent, the results of attaching to them the force of *assertion* are utterances that cannot be instantiations of the same illocutionary act, whereas the results of attaching to them the force of *question* are utterances that *can* be instantiations of exactly the same illocutionary act.[1]

---

1 The correct claim that relevant occurrences of *Is Sam autistic?* and *Isn't Sam autistic?* (or of *Is there nothing divisible by zero?* and *Is there something divisible by zero?*) can instantiate exactly the same *illocutionary* act is not, of course, to be confused with the incorrect claim that such occurrences are identical in all their *pre*-illocutionary or *post*-illocutionary properties. They are obviously different in their pre-illocutionary properties, since they constitute different *utterance* acts (in the sense of Searle 1969: 24); that difference, however, is not relevant to the present discussion, since, as Searle would be the first to acknowledge, utterance act identity is neither necessary nor sufficient for illocutionary act identity. And they may also be different in their post-illocutionary, and in particular in their perlocutionary, properties (in the sense of Searle 1969: 24), since their perlocutionary effects (including their 'preferred' conversational sequels) may be different; that difference, however, is not relevant either, since, as Searle would again acknowledge, perlocutionary effect identity is neither necessary nor sufficient for illocutionary

The implications of that fact on the issue under discussion here are straightforward: in order to make the fact consistent with the Content Invariance Thesis, one would have to abandon either the Propositional Distinctness Thesis or the Illocutionary Distinctness Thesis. Choosing the first option would allow one to preserve CIT along with IDT, but would require maintaining, contra PDT, that two propositions can be identical *even when they have nonidentical truth conditions*— which would mean that, for example, the proposition <u>that Sam is autistic</u> and the proposition <u>that Sam is not autistic</u>, despite their obvious truth conditional difference, can mysteriously become *the same proposition* when they are attached to the force of question, though they enjoy separate identities when they are attached to the force of assertion. Choosing the second option would allow one to preserve CIT along with PDT, but would require maintaining, contra IDT, that, although the attachment of its force to a particular proposition is somehow essential to an illocutionary act's identity, and although two propositions are nonidentical whenever they have nonidentical truth conditions, nevertheless the identity of an illocutionary act does *not* depend on whether the propositions to which its force is attached are or are not identical. Clearly, both options are deeply unsatisfactory, if they are intelligible at all. If, as the first option demands, two propositions are to be supposed to be identical even when the one is true and the other false with respect to the *same* situation, then it is completely unclear what sorts of entities propositions could possibly be, and they are certainly not the sorts of entities that anyone has ever taken them to be. And if, as the second option demands, two illocutionary acts are to be supposed to be identical *whether or not* the propositions to which they attach the same force are identical, then it is completely unclear what justification could possibly be given for insisting that the identity of *every* illocutionary act depends on the attachment of its force to a particular proposition.

The solution, I suggest, is to maintain PDT and IDT just with respect to those illocutionary acts (for example, positive and negative *assertions*) whose analysis mandates, or at least tolerates, the hypothesis that their content is propositional, and to deny that yes–no *questions* have contents that *are* propositional, and so to reject the Content Invariance Thesis—for, clearly, if the content of a yes–no question is *not* a proposition, then the yes–no question cannot consist in attaching a force to *the same proposition* as its grammatically corresponding assertion; and the idea that a yes–no question and its grammatically corresponding assertion

---

act identity. The only fact relevant to the present discussion is that the same *illocutionar*y act can be instantiated by the occurrences under discussion, even though the pre-illocutionary or post-illocutionary properties of those occurrences may be different.

do attach their respective illocutionary forces to *the same proposition* was precisely what the Content Invariance Thesis was all about. Restricting the domain of applicability of PDT and IDT in the indicated way would allow one to recognise the obvious fact that, for example, *Sam is autistic* and *Sam isn't autistic* instantiate *different* assertions rather than different ways of making the *same* assertion. Rejecting, at the same time, the Content Invariance Thesis would allow one to recognise the equally obvious fact that *Is Sam autistic?* and *Isn't Sam autistic?* do *not* instantiate different questions, but merely constitute different *ways* of posing the same question (constitute, as one might put it, different utterance acts dedicated to the performance of the same illocutionary act). Since both facts are central and undeniable, any theory of illocutionary acts that fails to accommodate them would be unsatisfactory. And Searle's theory does fail to accommodate them, because of commitments that, without proper motivation, it undertakes at a fundamental level: if one believes that the sententially representable content of *every* illocutionary act is necessarily a proposition, then, given that yes–no questions are obviously illocutionary acts with sententially representable contents, one is led to assume that their contents *must* be propositions; and if one accepts that the content of a yes–no question must be a proposition, then it is difficult to imagine what that proposition could possibly be unless it was the same as the proposition expressed by its grammatically corresponding assertion; the Content Invariance Thesis becomes then irresistible, with the unacceptable consequences we have noticed. But the unacceptable consequences do not follow if one refuses to grant the initial assumption, and that by itself is a perfectly good reason for refusing to grant it.

Notice that the initial assumption has, in any case, nothing self-recommendable about it, and that, at least as far as yes–no questions are concerned, definitely preferable alternatives to it can be made available. Certainly no *argument* has ever been given for the assumption that the sententially representable content of every illocutionary act *must* be a proposition; the assumption seems to have been mainly motivated by the thought that, since the contents of assertions can profitably be taken to be propositional (actually, on many traditional accounts, "proposition" just *means* "content of an assertion"), one must *hope* that the sententially representable contents of all other kinds of illocutionary act will turn out to be propositional as well; but the only consideration nurturing that hope appears to have been that, if the sententially representable content of every nonassertoric illocutionary act turned out to be propositional, then the principle of least effort (which is sometimes misidentified as "the principle of compositionality") would have been vindicated as a useful methodological principle in the

philosophy of language: one would have nothing more to do in analysing illocutionary acts than recycling a ready-made account of propositions and appending to it an account of force. As one should expect, however, the principle of least effort is not a reliable guide to truth, and the case of yes–no questions makes that especially clear, since the hypothesis that the content of yes–no questions is propositional leads, as we have seen, to consequences that no account of either propositions or illocutions could credibly endorse.

Besides, virtually no extant account of the content of yes–no questions in formal accounts of meaning (for a survey of such accounts, see Higginbotham 1996) has any use for the view that the content of a yes–no question is a proposition, suggesting instead (to put it in a way that abstracts away from differences among individual proposals, and adds an often neglected but essential condition) that its content is *the set of alternatives* among which some must be eliminated and some others must be selected if the question is to be *settled*. And these accounts, whatever their other limitations might be, can certainly avoid the problems that make the propositional account founder (even though they have not been put forward with these problems in mind), since, combined with the innocuous interpretation of the force-content distinction that I distinguished from the standard one, they can easily acknowledge the obvious identities and differences between yes–no questions that the standard interpretation is forced to deny—the fact, for example, that relevant occurrences of the interrogatives *Is Sam autistic?* and *Isn't Sam autistic?* instantiate the very same yes–no question, the fact that relevant occurrences of the interrogatives *Is there nothing divisible by zero?* and *Is there something divisible by zero?* instantiate the very same yes–no question, and the fact that the question that the first two of these interrogatives co-instantiate is not the same as the question that the last two co-instantiate. Assuming the innocuous interpretation of the force-content distinction, and taking the content of a yes–no question to be the set of answers that would be relevant to its settlement, the observed identities and differences would be accounted as follows: *Is Sam autistic?* and *Isn't Sam autistic?* instantiate the *same* yes–no question because, besides having the same illocutionary force, they have the same content; and they have the same content because each determines the same *set* of relevant answers (namely, the set {"Sam is autistic", "Sam isn't autistic"}), and each is settled by selecting the *same* member of that set to the exclusion of the other. Similarly, *Is there nothing divisible by zero?* and *Is there something divisible by zero?* instantiate the *same* question since, besides having the same illocutionary force, they have the same content; and they have the same content because each determines the same *set* of relevant answers (namely, the set {"There is nothing divisible by zero", "There is something divisible by zero"}), and each is settled

by selecting the *same* member of that set to the exclusion of the other. Finally, the question co-instantiated by *Is Sam autistic?* and *Isn't Sam autistic?* is *not* the same as the question co-instantiated by *Is there nothing divisible by zero?* and *Is there something divisible by zero?* because, although the illocutionary force of all four of these interrogatives is the same, the common content of the first two is different from the common content of the last two; and the common content of the first two is different from the common content of the last two because the answer set determined by each of the first two is disjoint from the answer set determined by each of the last two, with the consequence that no settlement of either of the first two can amount to a settlement of either of the last two, or conversely. Notice that what makes an account of this sort capable of acknowledging these obvious identities and differences between yes–no questions is precisely the fact that it takes the content of a yes–no question *not* to be a proposition: although an answer may have a truth condition, and although a proposition (supposing that answers regularly express propositions) must have a truth condition, a *set* of answers, or a *set* of propositions, cannot, as such, have a truth condition (just as a *set* of people cannot, as such, have a sex, even though each member of the set may well have a sex); and since an object that is incapable of having a truth condition cannot be a proposition, it follows that the object that, on such an account, provides the content of a yes–no question (namely, the answer *set* relative to which the question could be *settled*) cannot be a proposition.

There are, of course, significant issues that should be addressed before an account of this sort could be taken to constitute a full account of the content of yes–no questions. For example, it is usually taken for granted in formal accounts of yes–no questions that, although a question's answer set, and so a question's content, cannot be a proposition, the individual members of that set might well be taken to be propositions. But I think that this is a far from obviously correct assumption, and that a far better way of conceiving of a question's possible answers would be to conceive of them as possible *illocutionary acts* of certain types (a given yes–no question's answer set may consist of a possible assertion and a possible denial, a different yes–no question's answer set may consist of a possible permission and a possible prohibition, a still different yes–no question's answer set may consist of a possible approval and a possible disapproval; and so on). What is of primary importance in the present context, however, is the idea that, no matter how a yes–no question's possible answers are to be best characterised, it is the *set* of those answers, which cannot be a proposition, that provides the question's content. Since we already know that the assumption that the content of a yes–no question is a proposition has clearly unacceptable consequences, and since the alternative idea just rehearsed can be easily put to work

to account for at least those differences and identities between yes–no questions that a propositional conception of their content cannot coherently account for, the conclusion must be that there is no good reason to uphold either Searle's Content Invariance Thesis or the even more fundamental thesis that motivates it, namely, the thesis that *every* sententially representable illocutionary act content is a proposition. My purpose in the next section will be to reinforce that conclusion by providing independent evidence in its favour that derives from considering how Searle's propositional conception of the content of yes–no questions creates insuperable problems for his well-known and widely influential classification of illocutionary acts.

3 Yes–no questions and the classification of illocutionary acts

Searle's classification (Searle 1975a, reprinted as Chapter 1 of Searle 1979) explicitly assumes that all contentful illocutionary acts have the general form *F(p)* where *F* is a variable for forces and *p* a variable for *propositions* to which forces are attached, and, taking types of act to be determined by types of force, aims at arriving at a system of categories within which all possible types of force (and so, all possible types of illocutionary act) can be accommodated. The proposed fundamental principle of classification is the intrinsic point or purpose of an illocutionary act with a given force—in other words, the purpose that a speaker cannot fail to have if he is to perform an illocutionary act with that force—and, on its basis, Searle arrives at his five major categories of illocutionary acts (assertives, directives, commissives, expressives, and declarations) whose respective intrinsic points correspond to the five major types of purpose that, according to Searle, the expression of a proposition could possibly have. A speaker performs an assertive illocutionary act iff his point is to present as actual, without thereby making actual, the state of affairs represented by the proposition he expresses. A speaker performs a directive illocutionary act iff his point is to attempt to make his hearer make actual the state of affairs represented by the proposition he expresses. A speaker performs a commissive illocutionary act iff his point is to commit himself to making actual the state of affairs represented by the proposition he expresses. A speaker performs an expressive illocutionary act iff his point is to express his feelings and attitudes toward an already existing state of affairs represented by the proposition he expresses. And finally, a speaker performs a declarational illocutionary act iff his purpose is to make actual, simply by presenting

it as actual, the state of affairs represented by the proposition he expresses. "If we take the illocutionary act (that is, the full blown illocutionary act with its illocutionary force and propositional content) as the unit of analysis", Searle says in a later summary of his classification, "then we find there are five general ways of using language, five general categories of illocutionary acts. We tell people how things are (Assertives), we try to get them to do things (Directives), we commit ourselves to doing things (Commissives), we express our feelings and attitudes (Expressives), and we bring about changes in the world through our utterances (Declarations)" (Searle 1979: vii-viii); it may be worth noting that Searle's fundamental assumption that this classification is exhaustive is the basis of the later formalisation of his theory of speech acts in Searle and Vanderveken (1985).

Now, if, as Searle insists, these are the "five and only five" (1998: 148) categories of illocutionary acts that exist, one should expect that yes–no questions, which are undoubtedly central illocutionary acts, could be easily seen to belong to one or another of these five categories. However, it can easily be seen (though, to my knowledge, it has never been pointed out) that yes–no questions cannot belong to any of these categories, *if* their content is taken to be, as Searle's Content Invariance Thesis demands, the same as the content of their grammatically corresponding assertions. Take any yes–no question, for example the question *Does Sam smoke habitually?*, which, according to Searle, expresses the same proposition as its grammatically corresponding assertion—namely, the proposition <u>that Sam smokes habitually</u>. Clearly, a speaker who poses the question, *Does Sam smoke habitually?* does not aim to "present it as actual" that Sam smokes habitually, so he is not performing an assertive illocutionary act; but neither does he aim to "try to get his hearer to make it actual" that Sam smokes habitually, so he is not performing a directive illocutionary act; nor does he aim to "commit himself to making it actual" that Sam smokes habitually, so he is not performing a commissive illocutionary act; nor does he aim to "express his feelings and attitudes towards the antecedently obtaining fact" that Sam smokes habitually, so he is not performing an expressive illocutionary act; nor does he aim to "make it actual just by virtue of presenting it as actual" that Sam smokes habitually, so he is not performing a declarational illocutionary act. But if a speaker who, in posing the question, *Does Sam smoke habitually?*, does not attach to the proposition that, allegedly, he thereby expresses—that is, to the proposition <u>that Sam smokes habitually</u>—neither an assertive, nor a directive, nor a commissive, nor an expressive, nor a declarational illocutionary force, it follows that yes–no questions cannot belong to any of the "five and only five" categories of illocutionary acts that, according to Searle, exist. And since a classification of illocutionary acts that provides no place at all for yes–no questions is obviously unacceptable, it must be

concluded that Searle's Content Invariance Thesis has the significant added disadvantage of making it impossible for him to construct an acceptable classification of illocutionary acts.

As far as I can tell, Searle has never clearly recognised either this problem or the fact that, in order to properly respond to it, he should reconsider either his method of classifying illocutionary acts or his views about the content of yes–no questions (or, possibly, both). On the contrary, his position appears to be that his classification is perfectly satisfactory as it stands and that yes–no questions, with exactly the content prescribed for them by the Content Invariance Thesis, can be easily accommodated within one of the categories that the classification provides. When, however, one turns to the actual analysis of yes–no questions that Searle sketches in the course of arguing that they can be accommodated within his classification, what one finds is, first, that the proposed analysis, whether viable or not, ascribes to yes–no questions a content that is *not* the content prescribed by the Content Invariance Thesis, and second, that the proposed analysis is, in any case, not viable. I will now argue for these points in turn, and will conclude that there is no way of either properly analysing or properly classifying yes–no questions within Searle's system without rejecting the Content Invariance Thesis and the standard interpretation of the force-content distinction that underlies it.

Searle's view as to how questions in general, and yes–no questions in particular, could be accommodated within his proposed classification of illocutionary acts is stated very simply: "Questions", he says, "are a subclass of directives since they are attempts by S [that is, by the speaker] to get H [that is, the hearer] to answer, i.e. to perform a speech act" (1979: 14; interestingly, the same sort of account is suggested in Frege 1918). What Searle appears to be forgetting in putting forward this classificatory proposal, however, is that *if* yes–no questions are indeed directives by means of which speakers try to elicit answers from hearers (a claim that, just for the time being, we shall leave undisputed), then their propositional contents *cannot* be the same as the contents of their grammatically corresponding assertions, contrary to what the Content Invariance Thesis demands. For, by Searle's own definitions, each directive is an illocutionary act whose propositional content *specifies the action that the speaker is attempting to make the hearer perform* (if, for example, a speaker requests of his hearer to climb Mount Everest, then the propositional content of the request is <u>that the hearer climb Mount Everest</u>; if a speaker requests of his hearer to kill the discoverer of penicillin, then the propositional content of the request is <u>that the hearer kill the discoverer of penicillin</u>; and so on). Consequently, if yes–no questions are, as Searle contends, a particular subclass of directives where the action that the

speaker is attempting to make the hearer perform is the action of supplying an answer, then *that action itself* must be represented in the propositional content of those directives. And if it *is* represented there, then it is logically impossible for the propositional contents of those directives to ever be the same as the propositional contents that, according to Searle, yes–no questions allegedly *share* with their grammatically corresponding assertions. If, for example, the yes–no question *Is water odourless?* really is a request to the effect that its hearer tell its speaker whether water is odourless, then, obviously, its propositional content (whether explicitly given or not) is <u>that the hearer tell the speaker whether water is odourless</u>—in other words, it is the same as the propositional content that would be expressed by the same speaker if he was to address to the same hearer the imperative *Tell me whether water is odourless!*; but the propositional content <u>that the hearer tell the speaker whether water is odourless</u> is certainly not identical to the propositional content <u>that water is odourless,</u> which Searle was taking to be the propositional content of the question *Is water odourless?*, when he was claiming that that question has *the same propositional content* as the assertion *Water is odourless*; for, the proposition <u>that water is odourless</u> (which, of course, *is* the proposition expressed by the assertion *Water is odourless*) concerns only water and odourlessness, and not any speakers, any hearers, or any speech acts (of telling, answering, replying, or whatnot) that speakers may be attempting to get hearers to perform. Clearly, then, no analysis of yes–no questions could consistently maintain both that yes–no questions are directives aimed at eliciting answers *and* that their propositional contents are the same as the contents of their grammatically corresponding assertions. (Note that Frege [1918] gets himself involved in the same inconsistency as Searle in this connection.) And this, of course, means that, even if the analysis of yes–no questions implied by Searle's classificatory proposal were viable, it would still require him to fully reject the Content Invariance Thesis, according to which the propositional content of a yes–no question simply *is* the same as the propositional content of its grammatically corresponding assertion.

   Let me finally turn to the analysis of yes–no questions implied by Searle's classificatory proposal and briefly examine its central claim that a question as to whether something is the case is equivalent to a speaker's request that his hearer tell him whether that thing is the case. Notice first of all that even if the indicated equivalence did hold, it is quite unclear that it could be taken to provide an adequate *analysis* of yes–no questions, since, understood as an analysis, the proposed equivalence would seem not to satisfy the requirement of noncircularity that an adequate analysis should satisfy. For, arguably, a speaker tells a hearer *whether* something is the case (as opposed to telling the hearer *that* something is

the case) if and only if the speaker addresses to the hearer an utterance that settles the actual or potential *question* as to whether that thing is the case. And if this is so, the claim that the *question* whether something is the case is equivalent to a speaker's request that his hearer tell him *whether* that thing is the case would not constitute, even if it were otherwise acceptable, a noncircular analysis of questioning. In any event, the claim is not acceptable on several independent grounds. A first (and widely neglected) type of evidence against it comes from the fact that a question whether something is the case and a request for telling whether that thing is the case are two types of speech act that, on the one hand, may sometimes not even be grammatically interchangeable and, on the other hand, are often constrained to receive markedly divergent interpretations even when they are grammatically interchangeable. Thus, although there is nothing wrong about the request in (23), there is nothing right about the purported question in (24), even though the latter should be, if the proposal under discussion were correct, in all relevant respects indistinguishable from the former:

(23)   If you are interested in my offer, please tell me whether that is so.
(24)   * If you are interested in my offer, is that so?

What is more, even when both a yes–no question and its allegedly equivalent request are grammatical and interpretable, the interpretation of the one may be *required* to be different from the interpretation of the other. For example, (25) has obviously not the same meaning as (26), but this fact could hardly be explained if one accepted the idea that a speaker's question as to whether something is the case is the very same thing as a speaker's request that his hearer tell him whether that thing is the case:

(25)   If you are not busy, will you come to my party?
(26)   If you are not busy, tell me whether you will come to my party.

But the most direct evidence against the proposed analysis comes from two other facts. The first is that a speaker can without semantic oddity pose a question and explicitly deny that he is requesting of his hearer to provide its answer. For example, there is nothing semantically incongruous about utterances such as those in (27) and (28), but something should be radically incongruous about each one of them if, as Searle supposes, questions were *necessarily* requests aimed at eliciting hearer responses:

(27)   Will my lottery ticket win?—I am not, of course, asking you to tell me whether it will, I simply wonder whether it will.

(28)   Do I really want to marry her?—I am not, of course, asking you to tell me whether I do, I am just trying to make up my mind.

And the second fact is that questions in general, and yes–no questions in particular, can have perfectly meaningful occurrences in contexts where their speakers neither have nor can presume to have any hearers, and therefore cannot coherently be supposed to be requesting verbal actions on the part of their nonexistent hearers. For example, several minutes after departing alone from his house, a speaker might interrupt his solitary walk and utter the sentence in (29):

(29)   Have I locked the front door?

Since such a speaker could very well know that he does not have a hearer, he could hardly be supposed to be "requesting of his hearer to provide an answer." But he would have certainly *posed the question* whether he has locked his front door, and anyone knowing the language would know exactly what the *possible* answers would be, relative to which his question could be settled.

    Indeed, given the absence of a *necessary* connection between posing a question and requesting of a hearer to provide an answer, it could be suggested that the fact that, when one poses a question, one is *usually*, but not *invariably*, interpreted as requesting of a hearer to provide an answer, should be analysed by Searle along the same Gricean lines that he had followed (Searle 1975b, reprinted as Chapter 2 of Searle 1979) when discussing the familiar fact that, in uttering *any* sentence, one may often be interpreted as performing, apart from the direct illocutionary act that is part of the sentence's conventional meaning, various indirect illocutionary acts that are in no way part of the sentence's conventional meaning, but can reasonably be attributed to the speaker if there are no contextual indications disallowing the attribution. That these latter acts are *not* part of the sentence's conventional meaning, Searle was then claiming, is shown by the fact that they are *cancellable*—either in the sense that there exist contexts where the tendency to attribute them would not even arise, or in the sense that, even in contexts where the tendency to attribute them might arise, the speaker could effectively block their attribution by denying without semantic oddity the intention to perform them. What we have just seen is precisely that the interpretation of a question as a request for a hearer's answer is cancelable in both of these senses and should not, therefore, be taken by Searle to be part of the question's conventional meaning. And since Searle's intention in analysing questions as requests

for answers was, evidently, to elucidate their conventional meaning (to elucidate, in other words, the illocutionary acts that their speakers directly perform, and not any illocutionary acts that they might on occasion indirectly perform), it would seem that, if nothing else, Searle's own doctrine of direct and indirect illocutionary acts should have discouraged him from adopting the analysis under discussion.

I take it to be evident that, if there is no significant chance of accommodating yes–no questions (conceived of as a semantically homogeneous class) within Searle's category of directives, there is even less chance of accommodating them within any of his remaining categories of illocutionary acts. It seems, then, that there is no way, consistent with Searle's independent commitments, of making room for yes–no questions within his classification of illocutionary acts, and that the only basis on which he could maintain, together with those commitments, his fundamental claim that the classification is exhaustive would be the clearly untenable thesis that yes–no questions are not illocutionary acts at all. This result should not be surprising. If, as I have been arguing, yes–no questions have contents that are *not* propositional, then one should expect that a classification of illocutionary acts that, like Searle's, assumes that all such acts have contents that *are* propositional will not be able to accommodate yes–no questions. And the proper response to that situation is not, of course, to deny that yes–no questions are illocutionary acts, but rather to reject the standard interpretation of the force-content distinction that is uncritically applied to the analysis of all illocutionary acts, and to exploit the classificatory options that that rejection makes available. Assuming the innocuous interpretation of the force-content distinction that I distinguished from the standard one, a proper classification of illocutionary acts should, I suggest, accept as its fundamental distinction the distinction between *first-order* illocutionary acts, whose defining feature is that they attach their forces to contents that *are* propositions, and *higher-order* illocutionary acts, whose defining feature is that they attach their forces to contents that are *not* propositions but are, rather, sets of possible *first-order illocutionary acts*. Assertions and denials, acceptances and refusals, permissions and prohibitions, approvals and disapprovals are plausible examples of first-order illocutionary acts, since there is no obvious obstacle to assuming that the contents to which their respective forces are attached are propositions. However, questions of all sorts, and so yes–no questions, are higher-order illocutionary acts, since the contents to which their forces are attached are not propositions but rather sets of possible first-order illocutionary acts. A yes–no question, in particular, attaches its force to a set whose members are two possible first-order illocutionary acts (in many cases, a possible assertion and a possible denial, but in other cases, a possible

acceptance and a possible refusal, or a possible permission and a possible prohibition, or a possible approval and a possible disapproval, etc.) which are such that, *if* one of them were to be felicitously performed to the exclusion of the other, the question would have been settled.

This, of course, could only be the beginning of an adequate account of yes–no questions. For it is only by specifying exactly what is involved in a question's being *settled* by an appropriate choice from its answer set, and exactly what determines the appropriateness of such choices, that one would be in position to say exactly what the illocutionary *force* of questioning is; and it is only by analysing a question's illocutionary *force*, and not merely by describing the answer set that provides its content, that a full account of its meaning would be possible. Nevertheless, a proper conception of a question's content is an essential prerequisite to addressing the issue of its force, and my primary purpose in this essay has been to argue that that essential prerequisite is impossible to fulfil unless certain widely held assumptions about content in the theory of meaning are rejected. The most important of these assumptions is the assumption that the content (as distinct from the force) of every theoretically significant instance of meaningful speech consists in the expression of a proposition. If I am right, Searle has been too quick to concede to "the older philosophers" that propositions can be accorded that role, and to complain only that "their account was incomplete, for they did not discuss the different illocutionary acts in which a proposition could occur" (Searle 1969: 125). For if illocutionary acts as central as yes–no questions do *not* consist in the attachment of a force to a content that *is* propositional, then it seems that "the older philosophers", and their many contemporary followers, have made an even more limited contribution to the study of meaning than Searle was claiming they did when he made his own seminal contributions to that subject.

Chapter 12
Deontic trouble in speech act botany

1 Introduction

In the course of proposing his well-known classification of illocutionary acts into "five and only five" mutually exclusive and jointly exhaustive classes (assertives, directives, commissives, expressives, and declarations), John Searle notices a problem that acts of permission pose to him (Searle 1979: 14, 22): although he takes permissions to clearly belong to the same natural class as orders, commands, prohibitions, entreaties, etc., his classification of illocutionary acts does not seem capable of reflecting this fact, since permissions fail to exemplify what he takes to be the defining characteristic of the class of illocutionary acts to which he has allocated the others. Orders, commands, prohibitions, entreaties, etc. are, according to Searle, *directive* illocutionary acts, and the defining feature of these, he assumes (Searle 1979: 13), is that they are attempts by speakers to make hearers do things (or—one might add on Searle's behalf—to make hearers refrain from doing things). But, Searle observes, permissions cannot, given this definition of directives, be counted as directive illocutionary acts at all, since they are evidently not attempts by speakers to make hearers do things (nor—one might add on Searle's behalf—are they attempts by speakers to make hearers refrain from doing things). Therefore, the classification, as it stands, cannot reflect his strong pre-theoretical conviction that they belong to the same natural class as orders, commands, prohibitions, entreaties, etc.

Searle's response to this problem is not to reconsider his classificatory categories, but rather to define permission in terms of certain other uncontroversially directive acts and to thus conclude that, initial appearances notwithstanding, his classification does match his pre-theoretical convictions. The proposed definition has, at present, two variants, which are supposed to be equivalent by virtue of Searle's thesis that acts of forbidding that something be done are identical to acts of ordering that it not be done (Searle & Vanderveken 1985: 202). According to the first variant (Searle 1979: 22), permissions that something be done are nothing else but 'illocutionary negations' (also called 'illocutionary denegations') of orders that it not be done—in other words, the illocutionary act performed in uttering a sentence like "I give you permission to leave" is the same as the illocutionary act performed in uttering a sentence like "I do not order you not to leave". According to the second variant (Searle and Vanderveken 1985: 202), permissions

that something be done are nothing else but 'illocutionary negations' (also called 'illocutionary denegations') of prohibitions that it be done—in other words, the illocutionary act performed in uttering a sentence like "I give you permission to leave" is the same as the illocutionary act performed in uttering a sentence like "I do not forbid you to leave"). No matter which one of the variants you adopt, Searle believes, you have successfully reduced permissions to well-established species of directive illocutionary acts (notice the claim that "'permit' is the illocutionary denegation of 'forbid' *and thus* permission is directive", Searle & Vanderveken 1985: 195; emphasis added); you can therefore confidently regard the classification as reflecting your pre-theoretical conviction that permissions are illocutionary acts of the same kind as orders, commands, prohibitions, entreaties, etc.

I propose to show three things: first that, even if Searle's definitions were correct, they could not possibly resolve his classification problem; second, that they are not, in fact, correct; and third, that they are not equivalent to each other.

2 Problem A

Searle's thesis that his definitions of permission, assuming them to be correct, succeed in showing that this act is, after all, a directive illocutionary act (cf. the just quoted claim that "'permit' is the illocutionary denegation of 'forbid' *and thus* permission is directive") is very confused indeed. That thesis assumes that the operation of illocutionary negation on a speech act does not change the character of this act, and in particular that the negation of an act of ordering is itself an act of the same kind as an act of ordering, or that the negation of an act of forbidding is itself an act of the same kind as an act of forbidding. However, as Searle himself has pointed out elsewhere, "illocutionary negations in general change the character of the illocutionary act" (Searle 1969: 32), and this is especially clear when the illocutionary acts negated are acts of ordering or acts of forbidding. Someone saying

(1)   I do not order you not to leave.

or

(2)   I do not forbid you to leave.

can hardly be described as performing an act of the same kind as an act of order-ing, or an act of the same kind as an act of forbidding. He can only be described, to use terms that Searle himself had used in his general discussion of illocution-ary negation (Searle 1969: 32), as *refusing* to perform an act of ordering or an act of forbidding, or, alternatively, as *denying* that an act of ordering or an act of for-bidding is being performed. Therefore, *if* (1) and (2) are, as Searle contends, cor-rect translations of (3),

(3)   I give you permission to leave.

what Searle can claim he has achieved by devising them is not that he has shown that acts of permitting are a special sub-class of the class of acts to which acts of ordering or acts of forbidding belong (and that, therefore, they can confidently be regarded as directive illocutionary acts, given that orders and prohibitions are uncontroversially directive illocutionary acts), but rather that he has shown that acts of permitting are a special sub-class either of the class of acts to which acts of *refusing* belong or of the class of acts to which acts of *denying* belong. And since, as Searle well knows, refusals and denials are certainly *not* directive illo-cutionary acts (in Searle's own classification, refusals are unequivocally assigned to the class of *commissive* illocutionary acts, and denials are unequivocally as-signed to the class of *assertive* illocutionary acts; see Searle & Vanderveken 1985: 183, 195), he must conclude that his proposed definitions, *if* they are correct, achieve precisely the contrary of what he claims they achieve: instead of enabling him to maintain his view that permissions are, after all, directive illocutionary acts, they strictly disallow him to maintain that view.

3 Problem B

Let us now examine whether Searle's analysis of permission is correct in itself, since someone might be tempted to adopt it while ignoring the irresolvable prob-lems it creates for Searle's classificatory commitments. That the analysis is not, in fact, correct can be seen by noticing (among other things) that it is consistent neither with the normative prerequisites nor with the normative implications of acts of permission.

Regarding the normative prerequisites, notice that since issuing a permission (or an order, or a prohibition) requires that the person issuing the permission (or the order, or the prohibition) holds a position of authority that *entitles* him or her to the performance of these speech acts, a person could not coherently deny that

he or she is occupying such a position of authority while attempting to perform any of these speech acts. But now observe that, of the three utterances below, it is not (4) or (5), but only (6), that is incoherent:

(4)   Since I have no authority to order, permit, or forbid you anything, I do not forbid you to wear a tie.

(5)   Since I have no authority to order, permit, or forbid you anything, I do not order you not to wear a tie.

(6) * Since I have no authority to order, permit, or forbid you anything, I give you permission to wear a tie.

If Searle's analysis were correct, however, (4) and (5) should be just as incoherent as (6) is. And since they are not, it follows that, contrary to what that analysis claims, permitting that something be done is *not* the same thing as not forbidding that it be done, or not ordering that it not be done.

Regarding the normative implications, notice that if a speaker, who has the relevant entitlement, says to a hearer,

(7)   I give you permission either to leave my house or to marry my daughter.

he eliminates from the set of the options that he presents as normatively open to the hearer the option of the latter's staying in the speaker's house *without* marrying the speaker's daughter. In contrast, if a similarly entitled speaker addresses a hearer with either (8) or (9),

(8)   I do not order you either not to leave my house or not to marry my daughter.

(9)   I do not forbid you either to leave my house or to marry my daughter.

he does *not* eliminate from the set of the options that he presents as normatively open to the hearer the option of the latter's staying in the speaker's house without marrying the speaker's daughter. However, if the illocutionary act performed in uttering (7) was *the same* as the illocutionary act performed in uttering (8) or (9), as Searle's analysis entails, it would be impossible for the utterance of (7) to have normative implications that the utterance of (8) or of (9) does not have, and the contrast just noted would be non-existent. But since the contrast clearly exists, it follows that Searle's analysis cannot be maintained.

Finally, a different kind of reason against maintaining Searle's analysis of permission emerges when one notices that, as the incoherence of (10) or (11)

shows, a speaker cannot grant permission for a state of affairs that, given what he takes to be true, he ought to recognise as impossible:

(10) ⋆ Since you have never spoken to Mary, I give you permission to speak to her again.

(11) ⋆ Since you have never visited Boston, I give you permission to revisit it.

If Searle's analysis was correct, however, the following utterances should be just as incoherent as the preceding ones, since, on that analysis, their speakers would be attempting to perform exactly the same illocutionary acts that the speakers of the preceding ones cannot perform:

(12) Since you have never spoken to Mary, I do not order you not to speak to her again.

(13) Since you have never visited Boston, I do not order you not to revisit it.

But since (12) and (13), unlike (10) and (11), are perfectly coherent utterances, it follows that Searle's analysis cannot be maintained.

To sum up, Searle's analysis of permission is no more correct than his assumption that, if it *were* correct, it would allow him to solve his classification problem.

4 Problem C

Let us finally examine whether Searle was right in believing that his two definitions of permission are equivalent, since one might be inclined to accept his reason for regarding them as equivalent—in other words, his thesis that acts of forbidding that something be done are identical to acts of ordering that it not be done—, while acknowledging that neither of them offers an acceptable analysis of permission.

The definition of acts of forbidding that was supposed to support Searle's equivalence claim may seem obvious—"I forbid you to leave" and "I order you not to leave" can easily appear to be interchangeable—, but closer inspection of relevant evidence reveals it to be incorrect. For one thing, the definition implies that (14) and (15) can be used to perform exactly the same illocutionary act,

(14) I forbid your future children to stay in this country.

(15) ⋆ I order your future children not to stay in this country.

but this cannot be right, since (14) is perfectly meaningful whereas (15) does not even make sense. For another thing, the definition implies that (16) and (17) can be used to perform exactly the same illocutionary act,

(16)   I forbid those applying for this job to have worked for the government.

(17)   ⋆ I order those applying for this job not to have worked for the government.

but this cannot be right either, since (17) is nonsensical whereas (16) isn't. Finally, the definition implies that (18) and (19) can be used to perform exactly the same illocutionary act,

(18)   I never give orders to you, therefore I forbid you to give orders to me.

(19)   ⋆ I never give orders to you, therefore I order you not to give orders to me.

but this seems hardly possible, since (19) is incoherent whereas (18) is not.

In short, Searle's claim that his two definitions of permission are equivalent is no more successful than his claim that they are correct, or that they satisfactorily resolve his classification problem. And since it is Searle's definition of acts of forbidding that is responsible for the failure of the equivalence claim, it is not only his account of permission, but also his account of prohibition, that turns out to be unacceptable.

5 Conclusion

The general conclusion to be drawn is, it seems to me, this: If, as Searle and many others believe, a reasonable condition of adequacy on theories of speech acts is that they correctly analyse, and, in so doing, reflect the natural affinities between the concepts of ordering, prohibiting, and permitting, Searle's theory of speech acts fails to satisfy this condition.

Chapter 13
The gap between speech acts and mental states

1 Introduction

Suppose that, following John Searle, you have committed yourself to the thesis (call it "Thesis A") that three of the major categories of illocutionary acts are the categories of assertive, directive and commissive illocutionary acts, roughly characterisable as follows:[1]

(a) An *assertive* illocutionary act with propositional content p is an act whose speaker presents as actual the state of affairs represented by p. (Examples: asserting that p, claiming that p, predicting that p, informing someone that p, etc.)

(b) A *directive* illocutionary act with propositional content p is an act whose speaker attempts to make his hearer make actual the state of affairs represented by p. (Examples: asking someone to p, requesting someone to p, ordering someone to p, imploring someone to p, etc.)

(c) A *commissive* illocutionary act with propositional content p is an act whose speaker commits himself to make actual the state of affairs represented by p. (Examples: undertaking to p, promising to p, threatening to p, accepting to p, etc.)

Suppose further that, again following Searle, you have committed yourself to the thesis (call it "Thesis B") that the only kinds of things that are *intrinsically*, as opposed to *derivatively*, contentful are not linguistic acts like the act of asserting that something is the case, the act of requesting someone to make something the case, or the act of promising to make something the case, but rather *mental states* like the state of believing that something is the case, the state of desiring that

---

1 See Searle (1975a); Searle and Vanderveken (1985: 38-39, 54–56, 60–61, 99–101, 182–205).

someone make something the case, or the state of intending to make something the case.[2]

If you are committed both to Thesis A and to Thesis B, then you would be justified in fearing that a theory of speech acts that recognises the categories characterised in (a)–(c) as major illocutionary categories would be running the risk of being, at worst, misguided, and, at best, superficial, *unless* it could be shown that there are conceptual connections of some reasonably strong sort between its major categories of illocutionary acts and the major categories of an independently defensible taxonomy of mental states. Presumably for that reason, Searle has, in several of his recent writings, been eager to affirm that the following theses—among others of a similar sort—are all necessarily true:[3]

(i) For all Ss and all ps, if a speaker S performs an assertive illocutionary act with propositional content p, then S *expresses the belief* that p is the case.
(ii) For all Ss and all ps, if a speaker S performs a directive illocutionary act with propositional content p, then S *expresses the desire* that his/her hearer make p be the case.
(iii) For all Ss and all ps, if a speaker S performs a commissive illocutionary act with propositional content p, then S *expresses the intention* to make p the case.

So far as I know, Searle has offered only one kind of argument in favour of (i)–(iii) and of some other theses of the same kind—namely, that unless these theses were supposed true, one could not explain the oddity (marked by asterisks below) of utterances like the following, whose speakers appear to attempt both to perform assertive, directive or commissive illocutionary acts and to dissociate themselves from the expression of the mental states that, according to (i)–(iii), are necessarily connected with the performance of acts of those types.

(1) *   I assert, but I don't believe, that you are lazy.
(2) *   I request, but I don't want, your help.
(3) *   I promise, but I don't intend, to take you to the movies.

---

2 For Searle's views on the nature of mental states and their intrinsically contentful character, see Searle (1983). For important criticisms of these views—which I will be ignoring here—see Devitt (1990) and Dennett (1990). On the derivatively—as opposed to intrinsically—contentful character of linguistic acts, see especially Searle (1983: 4–13, 26–29, 160–179; 1986a; 1991a).
3 See Searle (1983: 9–10, 174–175, 177–179; 1986a; 1991a); Searle and Vanderveken (1985: 54–56, 102–105).

It is my purpose in this essay to suggest that this argument is unsuccessful and that, to this extent, Searle's project of integrating his philosophy of language into his philosophy of mind cannot be supposed to have been successful either.

2 The argument and its background

Before going any further, it will be important to realise that the conditions specified in (i)–(iii), though obviously related to, neither entail nor are entailed by, the so-called 'sincerity conditions' that Searle's theory of speech acts has, since its early formulation (in Searle 1969), always associated with illocutionary acts of the kinds characterised in (a)–(c): the condition requiring that a speaker performing an assertive illocutionary act should *express* the belief that a certain state of affairs obtains is not the same as the 'sincerity' condition requiring such a speaker to actually *have* the belief that that state of affairs obtains; the condition requiring that a speaker performing a directive illocutionary act should *express* the desire that his hearer make a certain state of affairs obtain is not the same as the 'sincerity' condition requiring such a speaker to actually *have* the desire that his hearer make that state of affairs obtain; and the condition requiring that a speaker performing a commissive illocutionary act should *express* the intention to make a certain state of affairs obtain is not the same as the 'sincerity' condition requiring such a speaker to actually *have* the intention to make that state of affairs obtain. The 'sincerity' conditions just mentioned should certainly be accepted as correct if understood in the right way, namely, as specifying types of mental states which, though not *necessarily* connected with illocutionary acts of the indicated types, are ascribed *by default* to speakers performing acts of those types—that is, are attributed to them as long as there is no evidence against their attribution, and can be consistently withdrawn whenever such evidence becomes available. And it is with this understanding that they have been proposed by Searle, who has always described them as conditions whose non-fulfilment entails not the non-performance but merely the 'defective' performance of assertive, directive and commissive illocutionary acts. However, precisely because the 'sincerity' conditions do not represent necessary features of the illocutionary acts in connection with which they have been proposed, they cannot, as they stand, fulfil Searle's larger, and more recent, project of *defining* his major categories of illocutionary acts by reference to the types of mental states that the conditions mention: since one can certainly assert things that one does not really believe, or request things that one does not really want, or promise to do things that one

does not really intend to do, it is impossible to correctly define assertions, requests or promises by reference to the *actual possession* of the relevant beliefs, desires or intentions. And it is at just this point that the introduction of additional conditions like those formulated in (i)–(iii) is supposed to prove its worth. For these conditions do purport to reveal necessary connections between types of illocutionary acts and types of mental states, not by saying, as the 'sincerity' conditions do, that one must *have* certain beliefs, desires or intentions in order to perform assertive, directive or commissive illocutionary acts, but only by saying that one must *express* such beliefs, desires or intentions in order to perform such acts—in other words, that one must *present oneself as having* those beliefs, desires or intentions, whether or not one *actually* has them. If, therefore, what these conditions purport to say is in fact true—and if, furthermore, one is prepared to grant that creatures that would be incapable of having genuine beliefs, desires or intentions could hardly *pretend* to have genuine beliefs, desires and intentions—the idea of a necessary connection between speech act concepts and mental state concepts would appear to have been vindicated.[4]

Searle's argument to the effect that the conditions specified in (i)–(iii) do in fact hold relies, as already noted, on the oddity of utterances like (1)–(3), and has been expressed many times, among which the following is representative:

> An insincere speech act is defective but not necessarily unsuccessful.…Nevertheless, successful performances of illocutionary acts necessarily involve the *expression* of the psychological state specified by the sincerity conditions

---

4 Confusingly, Searle sometimes, though not always, uses the term 'sincerity condition' to refer both to the possession and to the mere expression of mental states (see, for example, Searle and Vanderveken 1985: 22), though it is only the first usage, inaugurated in Searle (1969), that is clearly intelligible: a *sincerity* condition on a speech act is, presumably, a condition whose nonsatisfaction would make that act *insincere*, and what makes, for example, an assertion insincere is not that its speaker does not *express* a belief in the truth of its content, but rather that he does not *have* this expressed belief. (Indeed, since, on Searle's latest views, assertions are necessarily expressions of belief in their contents, the idea that the mere expression of such a belief would be sufficient to make any assertion sincere would have the absurd consequence that it is impossible for any assertion to ever be insincere—a consequence that Searle explicitly rejects.) The occasionally improper use of 'sincerity condition' to refer both to the (properly so-called) sincerity conditions of the earlier (1969) version of Searle's theory of speech acts and to conditions like (i)–(iii), which only appear in later versions, reflects, I suspect, Searle's unwillingness to clearly acknowledge the fact that the earlier version was, unlike the latter ones, quite unconcerned about the possibility of defining illocutionary acts in terms of mental states. For clear and interesting accounts of the nature and significance of this and related differences between the earlier and later versions, see Liedtke (1990) and Apel (1991). See also Harnish (1990).

of that type of act. The fact that the *expression* of the psychological state is internal to the performance of the act is shown by the fact that it is paradoxical to perform an illocution and to deny simultaneously that one has the corresponding psychological state. Thus, one cannot say 'I promise to come but I do not intend to come', 'I order you to leave but I don't want you to leave'…, etc. The reason for this is that when one performs the speech act one necessarily *expresses* the sincerity condition, and thus to conjoin the performance of the speech act with the denial of the sincerity condition would be to express and to deny the presence of one and the same psychological state.[5]

According to Searle, then, the reason for the oddity of (1) is that it is part of the *meaning* of "assert" that whoever asserts that *p* is the case expresses the belief that *p* is the case (in other words, presents himself *as* having the belief that *p* is the case), and that, therefore, a person who both asserts that something is the case and indicates that he does not believe it to be the case is a person who incoherently presents himself as both having and not having a certain belief. Similarly, the reason for the oddity of (2) is that it is part of the *meaning* of "request" that whoever requests that *p* be the case *expresses* the desire that *p* be the case (in other words, presents himself *as* having the desire that *p* be the case), and that, therefore, a person who both requests something and indicates that he does not want it, is a person who incoherently presents himself as both having and not having a certain desire. Finally, the reason for the oddity of (3) is that it is part of the *meaning* of "promise" that whoever promises to make *p* be the case *expresses* the intention to make *p* the case (in other words, presents himself *as* having the intention to make *p* the case), and that, therefore, a person who both promises to do something and indicates that he does not intend to do it is a person who incoherently presents himself as both having and not having a certain intention. In short, the explanation of the oddity of these and similar utterances rests, according to Searle, on the one hand, with the assumption that it is a *conceptual truth* that a person who asserts, requests or promises something presents himself *as* being in a certain mental state, and, on the other hand, with the assumption that it is incoherent to present oneself as simultaneously being and not being in a given mental state.

There are at least two ways in which one might establish that Searle's argument is unsuccessful: the first would consist in showing that the linguistic facts

---

5 Searle and Vanderveken (1985: 18–19; italics provided). See also Searle and Vanderveken (1985: 91); Searle (1983: 9–10, 1991b).

he appeals to do not *require* the kind of explanation he proposes, since an alternative explanation, which would *not* entitle him to the claim that there are necessary connections between kinds of speech acts and kinds of mental states, could be made available; the second would consist in showing that there are other linguistic facts that *exclude* the kind of explanation he proposes, and that can only be understood by assuming that there are *no* necessary connections between kinds of speech acts and kinds of mental states. We shall presently see that both lines of attack can be successfully deployed.

3 Why the argument fails

The first problem with Searle's argument is that the linguistic facts he appeals to do not require the semantic explanation he proposes, given that an alternative, pragmatic, explanation could be constructed for them, which, unlike the semantic one, would not allow the derivation of any conclusions, of the sort that Searle is interested in deriving, about necessary connections between types of illocutionary acts and types of mental states. The pragmatic explanation would begin by taking it as a matter of mutual knowledge between interlocutors that, since people are *usually* (though not invariably) expected to believe what they assert, to want what they request, and to intend to carry out what they promise to carry out, a *ceteris paribus* efficient strategy for *deceiving* others about one's beliefs, desires or intentions would be to assert what one does not believe, to request what one does not want, and to promise to do what one does not intend to do. The explanation would then appeal to a further piece of mutual knowledge between interlocutors, to the effect that a strategy to deceive others has, in general, little chance of being *successful* if the specific means chosen for implementing it make it *transparent* to others that one does have the intention to deceive them. And it would finally account for the oddity of (1), (2) and (3) by describing their speakers as persons who, on the one hand, give the appearance of having deceitful intents, and, on the other hand, appear to have chosen the ideal means for *failing* to fulfil those deceitful intents. The fact that (1) is odd, for example, would be explicable by reference, on the one hand, to the fact that someone who asserts something without believing it, is, in the absence of indications to the contrary, plausibly taken to be someone who wants to *deceive* his addressees about his real beliefs, and, on the other hand, to the fact that someone who wants to *successfully* deceive his addressees about his real beliefs should—unlike the speaker of (1)— *hide* from his addresses that he wants to deceive them about his real beliefs. Similarly, the oddity of (2) would be explicable by reference, on the one hand, to the

fact that someone who requests something without wanting it is, in the absence of indications to the contrary, plausibly taken to be someone who wants to *deceive* his addressees about his real desires, and, on the other hand, to the fact that someone who wants to *successfully* deceive his addressees about his real desires should—unlike the speaker of (2)—*hide* from his addresses that he wants to deceive them about his real desires. Finally, the oddity of (3) would be explicable by reference, on the one hand, to the fact that someone who promises to do something without intending to do it is, in the absence of indications to the contrary, plausibly taken to be someone who wants to *deceive* his addressees about his real intentions, and, on the other hand, to the fact that someone who wants to *successfully* deceive his addressees about his real intentions should—unlike the speaker of (3)—*hide* from his addresses that he wants to deceive them about his real intentions. In short, the oddity of (1), (2) and (3) would be explicable not by reference to the alleged fact that they violate a *semantic* constraint to the effect that assertions are necessarily expressions of beliefs, requests are necessarily expressions of desires, and promises are necessarily expressions of intentions, but rather to the fact that they violate an obvious *pragmatic* principle to the effect that if one wants to deceive others by asserting something that one does not believe, or by requesting something that one does not want, or by promising something that one does not intend to do, then it would be in one's own interest not to make it *manifest* to the intended victims of one's deception that one does not really believe what one asserts, that one does not really want what one requests, or that one does not really intend to do what one promises to do. And given this explanation, of course, the oddity of (1), (2) and (3) could hardly be used as evidence in favour of a necessary connection, of the sort Searle is interested in establishing, between expressions of mental states and performances of illocutionary acts.

The second—and most important—problem with Searle's argument is that, in addition to linguistic facts that do not *require* for their explanation the assumption of necessary connections between types of mental states and types of illocutionary acts, there are linguistic facts whose explanation is straightforwardly *incompatible* with that assumption. Notice that, since, by definition, you cannot turn a sentence that is semantically odd into a sentence that is not semantically odd *just* by changing the context in which the sentence is uttered, Searle's explanation predicts that utterances of sentences whereby a speaker attempts both to assert something and to deny that he believes it, or both to request something and to deny that he wants it, or both to promise something and to deny that he intends to do it, will be unacceptable in *all* types of context. But such sentences *are* perfectly acceptable in certain types of contexts (and these contexts are, not unexpectedly, precisely the ones for which acceptability rather than unaccept-

ability would be predicted on the pragmatic account, since they are contexts in which the utterance of the sentences would be unlikely or impossible to be construed as the manifestation of a *deception* plan).

Consider, to take the case of assertion first, the following utterances:

(4)   I don't believe I was there, but since the video shows that I was, I assert that I was.

(5)   I don't believe I have the virus, but since the experts are unanimous that I do, I assert that I do.

Given Searle's assumption that assertions are necessarily expressions of belief, what the speakers of these utterances do is, on the one hand, to present themselves as believing that something is the case, and, on the other hand, to present themselves as not believing that that thing is the case. But since it is incoherent to present oneself as simultaneously believing and not believing that something is the case, their utterances should be exactly as odd as the ones below:

(6) ⋆   I don't believe I was there, but since the video shows that I was, I believe that I was.

(7) ⋆   I don't believe I have the virus, but since the experts are unanimous that I do, I believe that I do.

The fact is, however, that, although (6) and (7) are indeed irreparably odd, (4) and (5) are not. This means that, contrary to what Searle contends, it is not part of the *meaning* of "assert" that whoever asserts that something is the case presents himself as *believing* that it is the case; and this in turn shows that it would not be merely needless but straightforwardly *wrong* to suppose that the speech act of assertion is *necessarily* an expression of the mental state of belief.[6]

Turning from assertions to requests (Searle's prime example of directive illocutionary acts) and to promises (his prime example of commissive illocutionary

---

6 An interesting consequence of the failure of Searle's claim that assertions are necessarily expressions of belief is that it not only deprives him of what he regards (Searle and Vanderveken 1985: 91; Searle 1991b: 185–188) as an 'explanation' of so-called "Moore's Paradox", but also makes it questionable whether such a thing as "Moore's Paradox"—interpreted as resulting from an allegedly *universal* incompatibility of assertions with expressions of disbelief—actually exists. I welcome this consequence—and am thus opposed both to Searle and to those of his critics (e.g. Malcolm 1991) who propose alternative explanations of the "paradox"—but will not further comment on this issue here.

acts), we find that utterances like those in (8) and (9) invite considerations analogous to the preceding ones:

(8)   I don't want you to do this dirty job, but since you are alone capable of doing it, and since it is tremendously important for the country, I request of you to do it.

(9)   Since I will die today, I don't intend to call you tomorrow–but since you mistakenly insist that I will be alive tomorrow, I promise to call you tomorrow.

Given Searle's view that requests are necessarily expressions of desires and that promises are necessarily expressions of intentions, it is difficult to imagine how these utterances could *avoid* being unacceptable. In fact, they should be exactly as unacceptable as the following utterances, in the first of which a speaker incoherently presents himself as simultaneously having and not having a certain desire, and in the second of which a speaker incoherently presents himself as simultaneously having and not having a certain intention:

(10)   * I don't want you to do this dirty job, but since you are alone capable of doing it, and since it is tremendously important for the country, I want you to do it.

(11)   * Since I will die today, I don't intend to call you tomorrow–but since you mistakenly insist that I will be alive tomorrow, I intend to call you tomorrow.

It is clear, however, that, though (10) and (11) are indeed unacceptable, (8) and (9) are not. This means that it is not part of the *meaning* of "request" that whoever requests something presents himself as wanting it, and that it is not part of the *meaning* of "promise" that whoever promises to do something presents himself as intending to do it. And this in turn shows that, contrary to what Searle wanted to establish, it is false that the speech act of requesting is necessarily an expression of the mental state of desiring, or that the speech act promising is necessarily an expression of the mental state of intending.

The moral should by now be clear: contrary to what theses (i)–(iii) state, the expression of the mental state of believing that p is not a necessary feature of assertive illocutionary acts whose propositional content is p; the expression of the mental state of desiring that p is not a necessary feature of directive illocutionary acts whose propositional content is p; and the expression of the mental state of intending that p is not a necessary feature of commissive illocutionary

acts whose propositional content is *p*. And if the expression of these mental states is not a necessary feature of the performance of these acts, then Searle's proposal to analyse the supposedly 'derived' contentfulness of illocutionary acts in terms of the supposedly 'intrinsic' contentfulness of mental states cannot be regarded as a successful proposal. We shall now proceed to a further, and independently interesting, confirmation of that failure.

4 Extending the counterarguments

The Searlian theses on assertive, directive and commissive illocutionary acts specified in (i)–(iii) presuppose, but are not presupposed by, the initial characterisations of these acts that Searle's theory provides under (a)–(c). Consequently, the failure of theses (i)–(iii) does not entail that the initial characterisations should themselves be rejected—it only entails that, even if they are correct as far as they go, they do not, by Searle's standards, go far enough, since the basic categories of speech act types that they establish do not correlate significantly with the basic categories of mental state types that, according to him, constitute the most fundamental form of contentfulness. There is, however, a category of speech act types whose members are *directly* characterised, within Searle's theory, as expressions of mental states of a particular kind; if, therefore, that characterisation is, as I think it is, open to objections of the same kind as those just raised against theses (i)–(iii), these objections would entitle us to conclude that Searle has not only failed to establish revealing correlations between an independently motivated taxonomy of basic speech act types and his preferred taxonomy of basic mental state types, but that he cannot even assume that an independently motivated taxonomy of basic speech act types actually exists. The category in question (known as the category of "expressive" illocutionary acts) is supposed to be the category of acts whose successful performance consists in the expression of emotional states of various kinds, and its prototypical members are acts of apologising, thanking and congratulating, which are supposedly definable by reference to the expression of feelings of regret, of gratitude and of gladness, respectively. Thus, apologies about a certain state of affairs *p* are, according to Searle, necessarily expressions of regret (whether real or feigned) about *p*, thanks about a certain state of affairs *p* are necessarily expressions of gratitude (whether real or feigned) about *p*, and congratulations about a certain state of

affairs *p* are necessarily expressions of gladness (whether real or feigned) about *p*.[7]

What is Searle's reason for thinking that these definitions are in fact correct? His reason, as one would expect, is that the supposition that they are correct provides the only possible explanation of the unacceptability of utterances like (12), (13) and (14), whose speakers would be described by him as incoherently presenting themselves as simultaneously having and not having a certain feeling of regret, a certain feeling of gratitude, or a certain feeling of gladness, respectively:

(12) ⋆ I apologise for insulting you, but I feel no regret about insulting you.
(13) ⋆ I thank you, but I do not feel grateful to you, for your help.
(14) ⋆ I congratulate you on, but I don't feel glad about, your promotion.

This claim, however, is, just like the analogous claims about assertive, directive and commissive illocutionary acts, open to two kinds of objection. The first objection is that, as far as the above evidence is concerned, Searle's semantic explanation is certainly not the only possible one, since an alternative, pragmatic, explanation could be offered: on that explanation, the reason why hearers find (12), (13) and (14) unacceptable would be that they assume, on the one hand, that a speaker who thanks without feeling gratitude, or who apologises without feeling regret, or who congratulates without feeling gladness, may plausibly be supposed, in the absence of indications to the contrary, to be a speaker who wants to deceive his audience about his real feelings, and, on the other hand, that a minimally intelligent speaker who wants to deceive his audience about his real feelings does not, as a rule, *reveal* to his audience that he does want to deceive them about his real feelings. The second—and most important—objection is that, apart from the fact that Searle's necessity claims are not *required* for making sense of the evidence he considers, there is independent evidence which shows that these claims are straightforwardly *incorrect*. Notice, to see this, that utterances like those in (15), (16) and (17) are perfectly coherent ones:

(15)   I apologise for doing such a painful thing to you; but since it was my only means of keeping you alive, I neither feel, nor can pretend that I feel, regret about doing it.
(16)   Since I asked you to do this, I thank you for doing it; but since I now realise how much harm it will cause me, I don't feel, nor can pretend that I feel, grateful to you for doing it.

---

7 See Searle (1975a); Searle and Vanderveken (1985: 39–40, 58–59, 62, 211–216).

(17)   Since you so desperately wanted to get your promotion, I congratulate you on getting it; but I don't feel, nor can pretend that I feel, glad that you got it, since I know that it will ruin your health.

Given Searle's definitions, however, these utterances should be just as incoherent as the ones below, since their speakers would be simultaneously presenting themselves as having and not having a certain feeling of regret, a certain feeling of gratitude, and a certain feeling of gladness, respectively:

(18) * I feel regret about doing such a painful thing to you; but since it was my only means of keeping you alive, I neither feel, nor can pretend that I feel, regret about doing it.

(19) * Since I asked you to do this, I feel grateful to you for doing it; but since I now realise how much harm it will cause me, I don't feel, nor can pretend that I feel, grateful to you for doing it.

(20) * Since you so desperately wanted to get your promotion, I feel glad that you got it; but I don't feel, nor can pretend that I feel, glad that you got it, since I know that it will ruin your health.

The fact, then, that it is only (18), (19) and (20), and not (15), (16) or (17), that are incoherent, shows that the meaning of "apologise", "thank", and "congratulate" is *not* such as to make it necessary for apologies to be expressions of feelings of regret, or necessary for thanks to be expressions of feelings of gratitude, or necessary for congratulations to be expressions of feelings of gladness. And since, without the assumption of such necessary connections, no definition *at all* of apologies, thanks and congratulations are deliverable by Searle's theory, it follows that his account of so-called "expressive" illocutionary acts creates even deeper problems for him than his account of assertive, directive and commissive illocutionary acts: not only can he not claim that there is a category of mental state types (namely, the category of emotional state types) with which an independently established category of "expressive" speech acts can be revealingly correlated, but he cannot even claim that there *exists* an independently established category of "expressive" speech act types, since the only feature that was supposed to hold together the putative members of the category—namely, that each of them is, *of necessity*, an expression of some emotional state—is not a feature satisfied even by its reputedly most characteristic members.[8]

---

8 For additional problems facing Searle's proposed definitions of so-called "expressive" illocutionary acts, see Tsohatzidis (1993b).

5 Conclusion

The conclusion we are now in a position to reach has a significant degree of generality. As is well known, there are, according to Searle, "five and only five" categories of illocutionary acts: assertives, directives, commissives, expressives and declarations.[9] The last mentioned among these categories may be bypassed in the context of the present discussion, since Searle has no consistent position as to whether declarations are or are not expressions of mental states.[10] Concerning the other four major categories, however, Searle is unequivocal about their alleged conceptual connections with specific categories of mental states (assertions being, in his view, necessarily expressions of beliefs; directives necessarily expressions of desires; commissives necessarily expressions of intentions; and expressives necessarily expressions of emotions), as well as about the supposed theoretical significance of this alleged fact—namely, that it substantiates the idea that the intentionality of linguistic acts is 'derivative' upon the 'intrinsic' intentionality of mental states. I have tried to show that Searle's linguistic argument for the existence of these necessary connections fails, and that, to this extent at least, the prospects of his grand project of showing that the intentionality of language is a by-product of the intentionality of mind are less bright than they might look.[11]

---

9 See, for example, Searle (1975a, 1986a); Searle and Vanderveken (1985: 37).

10 In some places (e.g. Searle 1975a: 360) he explicitly denies that declarations are expressions of mental states, whereas in other places (e.g. Searle 1983: 172; Searle and Vanderveken 1985: 57–58) he categorically affirms that they are.

11 An important contribution to the theory of speech acts that is as strongly committed as Searle's to the idea that illocutionary acts types are in part definable as expressions of mental state types is the one by Bach and Harnish (1979). Indeed, Harnish (1990) claims that the system of speech act analysis presented in Bach and Harnish (1979) serves Searle's purposes better than Searle's own system. And although that claim might, in some respects, be true, a look at Bach and Harnish's definitions of 'constatives' (Searle's assertives), directives, commissives, and 'acknowledgements' (Searle's expressives)—see Bach and Harnish (1979: 42, 44, 47–48, 49–50, 51)—will easily convince the reader that if the arguments here presented against Searle's definitions are sound, then they are just as forcefully applicable to the relevant parts of Bach and Harnish's definitions as well.

Chapter 14
A purported refutation of some theories of assertion

1 Preamble

Assertion is a fundamental type of speech act, and if there were reasons for think-
ing that a large number of influential philosophical analyses of it are mistaken, it
would be important to know what these reasons are. In an essay entitled, "Is as-
sertion social?" (Pagin 2004), Peter Pagin has claimed that a large number of in-
fluential philosophical analyses of assertion that, in his view, are united by the
fact that they represent assertion "as a social act" are fundamentally mistaken.[1]
My purpose in what follows is to argue that Pagin's defence of that claim is seri-
ously flawed, even granting some controversial assumptions on which it rests.

2 Pagin's argument

"In saying that a speech act type is *social*", Pagin states, "I mean that all acts of
this type not only have a social significance, consisting in the social effect in-
tended by the speaker, but also *communicate* this intended effect" (Pagin 2004:
834). Pagin immediately acknowledges that "it is not easy to specify" a "precise
technical sense" (2004: 835) for the notion of "what is communicated" by an ut-
terance, and proposes to explicate his use of that notion as follows: "What I in-
clude under 'what is communicated' is what anyone must know in order to un-
derstand the utterance" (Pagin 2004: 835). A speech act type, then, is "social" in
Pagin's sense when all its tokens are utterances that not only are intended to have
"social effects", but are also such that, in order to understand them, one must
know that they are intended to have those "social effects". (Exactly what it takes
for an effect to be "social" is an important question that Pagin does not address,
and some of the analyses he targets posit effects that would appear to be "social"

---

only in a very weak sense of that term.[2] However, at least some of the analyses he wants to dispute—and a consideration of these will be enough for present purposes—can, I believe, legitimately be regarded as being "social" in a robust sense of the term, since the effects they take assertions to be intended to generate are rights and obligations, and since rights and obligations are not the sorts of things that non-social beings could have.)

Having explained that a speech act is "social" in his sense when it not only is intended to have "social effects" but also cannot be understood unless it is known to be intended to have those "social effects", Pagin proceeds to the formulation of his central claim, which is that, although the vast majority of speech act types are, in the indicated sense, "social", assertion is "essentially different from virtually all other types" (Pagin 2004: 834) by virtue of not being "social". And since, in Pagin's view, there is a "prevalent conception" (Pagin 2004: 834) of the speech act of assertion according to which it *is* "social", it follows, Pagin concludes, that the prevalent conception in question is "deeply mistaken" (Pagin 2004: 834).

The central worry that Pagin expresses about the prevalent conception and that his argument (to be considered below) is meant to "make clearer as well as substantiate" (Pagin 2004: 836–837) is the following:

> [A] statement of the intended (or real) social effect of an assertion does not entail what is asserted, nor does what is asserted entail the statement of effect. Because of this, the significance of an assertion isn't exhausted by its *social* significance. Any description of the social significance will leave out the core of assertion, that it articulates a judgement about the world. It will always be possible to satisfy the description without expressing the judgment, i.e. without making an assertion. Because of this, no characterisation of the social significance of assertion, however correct and complete, can adequately define what assertion is. (Pagin 2004: 836)

Pagin's way of substantiating this central worry consists in the presentation of a "method for generating refutations" (Pagin 2004: 837) of analyses of assertion, which he regards as successfully applicable to all analyses of assertion that he

---

2 For example, Pagin construes Michael Dummett (Dummett 1973) as propounding an analysis of assertion according to which "to assert that *p* is to deliberately convey the impression of saying that *p* with the overriding intention of saying something true" (Pagin 2004: 845), and then claims that Dummett's analysis "should be counted as social, since it is based on the idea of conveying an impression" (Pagin 2004: 845).

takes to be paradigmatically "social". To see how the method is supposed to work, it is sufficient to consider two simple examples of Pagin's purported refutations, the first applying to analyses of assertion in terms of commitment, and the second to analyses of assertion in terms of entitlement and commitment.[3]

According to Pagin, "a popular idea that a number of philosophers have subscribed to, in one form or another, is that making an assertion essentially is to *commit* oneself to the truth of what is expressed" (Pagin 2004: 838; Pagin does not name particular philosophers as holders of that Peircian view, but he is, I believe, correct in claiming that the view is popular; one of its well-known modern exponents was William Alston[4]). Pagin represents the commitment analysis of assertion as follows (2004: 838),

(A1)   to assert that p is to commit oneself to the truth of p,

and claims that the basic question that one must answer in order to determine whether the analysis is correct is the following: "Is it the case that whenever I commit myself to the truth of some proposition, I also assert it?" (Pagin 2004: 838). The answer to that question, Pagin then contends, must be negative, since a speaker who utters an explicit performative utterance of the form in (1),

(1)   I hereby commit myself to the truth of p,

can be held to be performing the act of committing himself to the truth of p, but *cannot* be held to be performing the act of asserting that p. And the reason why, according to Pagin, a speaker of (1) cannot be held to be performing the act of asserting that p is that (1) can be true even though p is false: In uttering an "explicit performative formula" (2004: 838) of the form in (1), Pagin says,

> I *do* commit myself to the truth of p. But I don't assert that p. When I assert that p, what I say implies that p; it is logically incompatible with the falsity of p. By contrast, in the case of [(1)], what I say *is* logically compatible with the falsity of p. Hence, the act I perform by way of uttering [(1)] should be

---

3 These are also, in my view, the only kinds of analyses, among those that Pagin attacks, which could be called "social" in a robust sense of the term. This is so because, in the senses of "commitment" and "entitlement" intended by the analyses in question, having a commitment entails having an obligation, having an entitlement entails having a right, and neither obligations nor rights are things that non-social entities could be supposed to have.

4 Alston (1964). Alston's subsequent elaborations of the commitment view can be found in Alston (1994, 2000, and 2007).

regarded as committing me to the truth of p, but not as asserting that p. (Pagin 2004: 839)

It follows, Pagin concludes, that the analysis of assertion represented in (A1) is mistaken.

A similar argument is then advanced against Robert Brandom's analysis of assertion in terms of entitlement and commitment.[5] "According to Brandom", Pagin states, "the nature of assertion consists in the fact that the speaker achieves two different normative results at the same time: on the one hand he authorises the audience to claim anything that follows from what is asserted and on the other he undertakes the responsibility of justifying it" (Pagin 2004: 839). Pagin represents the Brandomian analysis as follows (2004: 839):

(A2)   to assert that p is to authorise the audience to claim whatever follows from p and to undertake the responsibility of justifying p.

That analysis too, Pagin then contends, must be rejected, since a speaker who utters an explicit performative utterance of the form in (2),

(2)   I hereby authorise you to claim whatever follows from p and I undertake the responsibility of justifying p,

can be held to be performing the acts of authorising and undertaking required by Brandom's analysis of the act of asserting a proposition, but *cannot* be held to be performing the act of asserting that proposition. And the reason why, according to Pagin, a speaker of (2) cannot be held to be performing the act of asserting that proposition is that the proposition in question can be false even though (2) is true: In uttering an explicit performative of the form in (2), Pagin says,

> I can indeed authorise you to claim whatever follows from p and undertake the responsibility of justifying p. Still, as little as before have I asserted that p. What I say is fully compatible with the falsity of p. So Brandom's analysis isn't acceptable. (Pagin 2004: 840)

Having proposed counterexamples of the same sort to several other analyses of assertion that he counts as "social", Pagin conjectures that no such analysis

---

5 Brandom (1994). For earlier and later presentations of that analysis see, respectively, Brandom (1983) and Brandom (2008).

can escape his "method of refutation" (Pagin 2004: 847), which may be summarised as involving, in its basic form, the following three steps (cf. Pagin 2004: 847–848):

> *Step 1.* Given an analysis of assertion of the form in (A), construct an explicit performative utterance of the form in (U) that instantiates the act type mentioned in the analysans of (A):
>> (A) to assert that *p* is to *X(p)*
>> (U) I hereby *X(p)*
> *Step 2.* Claim—on the basis of the observed compatibility of the truth of "I hereby *X(p)*" with the falsity of *p*—that a speaker who utters an instance of (U) "hasn't asserted that *p*" even though he has "manage[d] to *X(p)*."
> *Step 3.* Conclude that the analysis in (A) is false.

It will be noticed that Pagin's argument requires two assumptions that have been far from uncontroversial in the literature—namely, the assumption that explicit performatives are truth-apt, and the related but distinct assumption that they are descriptions of the acts their utterers perform. Thus, Pagin's claim that explicit performatives of the forms in (1) and (2) can be true even though *p* is false—a claim that, as we have seen, constitutes his reason for denying that (1) and (2) can be assertions whose content is *p*—, presupposes the thesis that explicit performatives (and not just the propositional clauses that may function as syntactic complements of their main verbs) can be bearers of truth values. And anyone who, like Austin (1962), rejects that thesis would be likely to regard Pagin's argument as misconceived from the start. In what follows, I will not challenge these controversial assumptions about performativity that Pagin's argument makes. But this is simply because I believe that the argument can be shown to fail even when these controversial assumptions are left unchallenged, and not because I believe that they are unchallengeable.[6]

---

6 For critical comments on other aspects of Pagin's position that I shall not be concerned with here, see Jary (2010: 58–59) and, especially, MacFarlane (2011: 92–93).

3 Pagin's mistakes, Part I

Recall that the central worry that Pagin's "method of refutation" was supposed to substantiate is that assertion cannot be adequately analysed in terms of its "social significance" because

> any description of the social significance will leave out the core of assertion, that it articulates a judgement about the world. It will always be possible to satisfy the description without *expressing the judgment, i.e. without making an assertion*. (Pagin 2004: 836; emphasis added)

It is evident that Pagin's method could not substantiate that worry if it delivered negative results when applied to an analysis of assertion that would encapsulate what, for Pagin, *is* the "core" of assertion—if, in other words, one was constrained to conclude, by applying Pagin's method, that it is *false* that to assert that p is to express the judgment that p. For, if something is false, no analysis can properly be criticised on the grounds that it *fails* to entail it. Therefore, if, by applying Pagin's method, one was constrained to conclude that it is *false* that to assert that p is to express the judgment that p, Pagin's method could not be appealed to in order to criticise what Pagin regards as "social" analyses of assertion on the grounds that they fail to entail that falsehood.

It is easy to see that acceptance of Pagin's method does dictate the conclusion that it is *false* that to assert that p is to express the judgment that p. Recall that, in order to derive the conclusion that some proposed analysis of the form "to assert that p is to $X(p)$" is false, the crucial claim that Pagin's method requires is the claim that a speaker who utters an explicit performative of the form "I hereby $X(p)$", and thus instantiates the act type mentioned in the analysans of the analysis, is *not* performing the act of asserting that p (even though he is performing the act of X-ing that p). Now notice that, according to Pagin's exposition of his method, what *justifies* that crucial claim whenever it is successfully made against a theory of assertion is simply this: that the propositions expressed by p and by "I hereby $X(p)$" are logically independent, in the sense that what is said by p neither entails nor is entailed by what is said by "I hereby $X(p)$." Thus, as we have seen, the key fact that, according to Pagin, justifies the claim that, when I say something of the form,

(1) I hereby commit myself to the truth of p,

I am "not asserting that p" (Pagin 2004: 839), is that "what I say is logically compatible with the falsity of p" (Pagin 2004: 839). And similarly, the key fact that, according to Pagin, justifies the claim that, when I say something of the form,

(2)   I hereby authorise you to claim whatever follows from p and I undertake the responsibility of justifying p,

I am not asserting that p, is that "what I say is fully compatible with the falsity of p" (Pagin 2004: 840).

The problem, however, is that if, as Pagin's method assumes, the compatibility of the truth of "I hereby $X(p)$" with the falsity of p is what justifies the claim that someone who says "I hereby $X(p)$" is not asserting that p, then exactly the *same* kind of justification is available for the claim that someone who utters an explicit performative of the form in (3),

(3)   I hereby express the judgment that p,

is *not* asserting that p, either, since the truth of "I hereby express the judgment that p" is obviously compatible with the falsity of p. (It can, for example, be false that there is a mouse in your house, but true that, by virtue of uttering the explicit performative, "I hereby express the judgment that there is a mouse in your house", I do express the judgment that there is a mouse in your house.) If this is so, however, Pagin's method cannot avoid deriving the conclusion that an analysis of assertion that would encapsulate what, for Pagin, *is* the very "core" of assertion, fails *for exactly the same kind of reason* as the reason that was supposed to condemn the "social" analyses of assertion that Pagin was targeting: Just as the method was constrained to interpret the compatibility of the truth of (1) or of (2) with the falsity of p as evidence that the "social" analyses of assertion represented in (A1) or (A2) are false,

(A1)   to assert that p is to commit oneself to the truth of p
(A2)   to assert that p is to authorise the audience to claim whatever follows from p and to undertake the responsibility of justifying p

it is *also* constrained to interpret the compatibility of the truth of (3) with the falsity of p as evidence that the (by hypothesis, *non-*"social") analysis of assertion represented in (A3) is false as well:

(A3)   to assert that p is to express the judgment that p.

And since that last analysis faithfully represents the alleged "core" of assertion that, according to Pagin, *any* adequate analysis of assertion ought to preserve, the fact that the analysis is *false* if Pagin's "method of refutation" is to be accepted as reliable, shows that Pagin cannot coherently appeal to his method in order to substantiate his advertised principal worry about the analyses of assertion he was aiming to undermine: No analysis can be faulted on the grounds that it fails to entail a falsehood, and so none of the "social" analyses of assertion that Pagin was aiming to undermine could be faulted on the grounds that it fails to entail what Pagin's method itself would be bound to regard as *false*.

4 Pagin's mistakes, Part II

At this point, one might suppose that the proper thing to do would be to *accept* Pagin's method as reliable and to conclude that Pagin's error simply consisted in the fact that he had misjudged its scope: although he had devised the method with the goal of undermining "social" and of supporting non-"social" analyses of assertion, what the method actually achieves, contrary to his intentions, is to show that *both* "social" and non-"social" analyses of assertion are wrong; and this—one might suppose—is a remarkable enough achievement, even if recognising it requires recognising that the specific goal that Pagin was aiming to attain by applying his method was unattainable.

I do not, however, believe that Pagin's pursuit of his goals could be credited with generating unintended but remarkable consequences. For, as I shall now argue, there is good reason to suppose that Pagin's "method of refutation" is *not* reliable, and therefore good reason to suppose that no theory of assertion, whether favoured or disfavoured by Pagin, would need to pay attention to its verdicts.

Any account of assertion should be in a position to distinguish between, on the one hand, *what it is* that a speaker is asserting and, on the other hand, the fact that the speaker is *asserting* it. And the ultimate basis of that distinction is simply this: that propositions of the forms *p* and *S asserts that p* are logically independent, in the sense that something's *not being* the case is consistent with its *being asserted* to be the case, and something's *not being asserted* to be the case is consistent with its *being* the case. Now, given that its capacity to recognise the logical independence of *p* and *S asserts that p* is a condition of adequacy of *any* analysis of the act of assertion, it follows that no criterion aiming to adjudicate

between rival analyses of that act will be successful if it stipulates that an adequate analysis should, on the contrary, *not* allow that *p* and *S asserts that p* be logically independent.

But that is just the criterion that Pagin's "method of refutation" in effect employs. For, what the method in effect instructs one to do is to reject any proposed analysis—be it "social" or non-"social"—of "S asserts that *p*" if the analysans of that analysis turns out to *be* logically independent of *p*. Thus, "S commits himself to the truth of *p*" cannot, according to Pagin's method, be the correct analysis of "S asserts that *p*", because *p* can be false even when "S commits himself to the truth of *p*" is true—in other words, because something's *not being* the case would be consistent, on that analysis, with what, according to the analysis, would amount to someone's *asserting it to be* the case. And similarly, as we have seen, "S expresses the judgment that *p*" cannot, according to Pagin's method, be the correct analysis of "S asserts that *p*", because *p* can be false even when "S expresses the judgment that *p*" is true—in other words, because something's *not being* the case would be consistent, on that analysis too, with what, according to the analysis, would amount to someone's *asserting it to be* the case.

What these verdicts evidently assume is that the question *whether a speaker has asserted something* could not be correctly answered in the affirmative unless its affirmative answer entailed that *what the speaker has asserted* is true. But since accepting this assumption is tantamount to accepting the obvious falsehood that it is impossible for erroneous or insincere assertions to exist (in other words, that it is impossible for a speaker who is *making an assertion* to be in error or to be lying in *what he asserts*), it is clear that no adequate theory of assertion could accept it. Therefore, no theory of assertion that is minimally adequate should be concerned about what Pagin's method demands. In particular, neither the "social" theories of assertion that Pagin was aiming to oppose nor the non-"social" theory of assertion that he presents himself as favouring should be concerned about what his method demands. For, despite their important differences, none of these theories are constitutionally incapable of distinguishing *what* is being asserted from the fact that it *is* being asserted, none of them are thus in danger of confusing the question whether a speaker is asserting something with the question whether what a speaker is asserting is true, and none of them are consequently committed to mistakenly supposing that it is impossible for a speaker who is *making* an assertion to be in error or to be lying in *what* he asserts.

5 Conclusion

Pagin's recipe for "refuting" theories of assertion cannot be recommended, whether in general or in relation to his stated goals. Even on the supposition that the recipe is reliable, it could not be used, in the way it was meant to be used, either against the kinds of theories Pagin was aiming to undermine or in favour of the kind of theory he was aiming to support. What is more, it cannot be supposed to *be* reliable, since its operation presupposes the denial of fundamental distinctions that any adequate analysis of assertion should observe. There are many interesting questions that rival philosophical analyses of assertion raise, but none of them, I submit, can be usefully addressed, let alone resolved, with the help of Pagin's recipe.

Chapter 15
Two consequences of hinting

1 Introduction

David Holdcroft has maintained[1] that John Searle's claim (Searle 1969: 19–21) to the effect that every illocutionary act performed inexplicitly can, in principle, be performed explicitly—a claim that is sometimes referred to as "the principle of expressibility"—is incorrect, since there is at least one illocutionary act, hinting, that cannot be correctly performed unless the person performing it says *less* than he means to say.

To date, Holdcroft's conclusion has been resisted by two philosophers, G. J. Warnock and (not unexpectedly) Searle himself, on the grounds that there are good reasons (which, interestingly enough, are not the same for Warnock and Searle) *against* counting hinting as an illocutionary act, and therefore no reason to think that it represents a threat to the principle of expressibility.

My purpose in this essay is twofold: first, to argue that neither Searle nor Warnock succeed in establishing their respective points, and that, to this extent, Holdcroft's claim that hinting represents a threat to the principle of expressibility retains its force; second, to argue that hinting creates a further (and somewhat different) problem for another component of Searle's philosophy of language, namely, for his descriptivist account of proper names.

2 Hinting and the limits of expressibility

Warnock's reason for doubting that hinting is an illocutionary act (Warnock 1983: 123–128) is that hinting (together with some other speech acts of apparently the same kind) does not satisfy a condition that, Warnock believes, is satisfied by all uncontroversial illocutionary acts. According to Warnock, when an uncontroversial illocutionary act is named under *y* in sentences of the form *In saying "X," I was y-ing* (where *X* stands for any non-performative utterance capable of being

---

1 Holdcroft (1978: 61–63). See also Holdcroft (1976).

used for the performance of an act of *y*-ing), no further question arises as to the specific content of this act of *y*-ing. For example, if someone says,

(1)   In saying "You will get your hair cut", I was giving an order.

it would be superfluous to ask *what* precise order has been given; obviously, the order given was that the interlocutor get his hair cut. However, Warnock contends, when an act like hinting is named under *y* in sentences of the form *In saying "X", I was y-ing*, a further question always arises as to the specific content of this act. For example, if someone says,

(2)   In saying "Did you meet Leo in Lyon?", I was giving a hint.

it would *not* be at all superfluous to ask of what precisely the hint in question consisted. Therefore, Warnock concludes, hinting is not an illocutionary act, and Holdcroft's argument against Searle, which presupposes that it is, has misfired.

It is unclear exactly how this conclusion would follow even if the contrast Warnock wished to emphasise was genuine, but this is of less importance than the fact that the contrast itself is spurious. If someone utters a sentence like,

(3)   In saying "The door is open, Walter", I was giving an order.

it would *not* be at all superfluous to ask what precise order he has given; he might be ordering Walter to close the door, he might be ordering Walter to get out, he might be ordering Walter to bring the guests in, and so on. But this is contrary to what Warnock should expect, given that orders are uncontroversially illocutionary acts. On the other hand, if someone utters a sentence like,

(4)   In saying "The number you are looking for is a multiple of 7", I was giving a hint.

it would *be* superfluous to ask what precisely the hint in question was: the hint was, plainly, that the number the speaker's interlocutor was looking for is a multiple of 7. But this is, again, contrary to what Warnock should expect, given that hinting is, for him, a doubtfully illocutionary act. Warnock, then, fails to establish a distinction between uncontroversially illocutionary acts and acts like hinting, and to thereby question Holdcroft's argument against Searle's principle of expressibility.

Searle's own attempt (Searle 1979: ix) to neutralise Holdcroft's argument is even more simple than Warnock's: hinting, he says, far from being an illocutionary act itself, is "a deliberately inexplicit manner of performing" a genuine illocutionary act; and since Holdcroft mistakes it for an illocutionary act, his argument must fail. Searle's attempt, however, is no more successful than Warnock's.

Suppose that Paul asks Peter to give him a hint in the direction of the correct answer to the question of,

(5)   Is George immodest?

and that Peter believes that the correct answer is,

(6)   George is immodest.

Now, if Peter is to say,

(7)   Well, George is not exactly modest,

he can certainly be described as asserting, 'in a deliberately inexplicit manner,' that George is immodest. But, contrary to what Searle should expect, he can hardly be described as giving the *hint* he was asked to give: saying, in this context, that George is not exactly modest, is not giving a *hint* in the direction of the correct answer; it is giving the answer itself. Suppose, on the other hand, that Paul asks Peter to give him a hint in the direction of the correct answer to the question of,

(8)   Who was the composer of *La Clemenza di Scipione*?

and that Peter believes that the correct answer to this question is,

(9)   Johann Christian Bach was the composer of *La Clemenza di Scipione*.

Now, if Peter is to say,

(10)   Johann Sebastian was not the only musician in the Bach family.

he can certainly be described as giving a hint in the direction of the correct answer to the question. But, contrary to what Searle should expect, he can hardly

be described as asserting, 'in a deliberately inexplicit manner', that Johann Christian Bach was the composer of *La Clemenza di Scipione*. Saying that Johann Sebastian was not the only musician in the Bach family is *not* saying, 'in a deliberately inexplicit matter', that Johann Christian Bach was the composer of *La Clemenza di Scipione*. It is simply *not* saying that thing at all. Searle, then, is no more successful than Warnock in making sense of the idea that hinting is not a genuine illocutionary act, and in thereby questioning Holdcroft's argument against the principle of expressibility.

In fact, Searle is not even consistent in his answer to the question of whether or not he takes hinting to be an illocutionary act. In the course of attacking Holdcroft, he says, as we have just seen, that it is not. But in a book he co-authored with Daniel Vanderveken, when explaining why certain verbs denoting illocutionary acts do not have performative uses, he writes,

> ...boasting, like hinting and insinuating, *is one of those illocutionary acts* that are intrinsically concealed. To the extent that the speaker makes his intentions explicit, as the performative verb would serve to do, to that extent his speech act would be less a case of hinting, insinuating, or boasting. (Searle and Vanderveken 1985: 189; emphasis added)

This not only means that, for Searle, hinting *is* an illocutionary act, but that the illocutionary act of hinting has, according to Searle, precisely the features that, on Holdcroft's view, make it a counterexample to the principle of expressibility: it is, Searle now says, 'intrinsically concealed', in the sense that one cannot both perform it and explicitly indicate, as the principle of expressibility would require, that one performs it. It seems, then, that, by its own inventor's lights at least, the principle of expressibility should not be maintained.[2]

3 Hinting and the limits of descriptivism

It is not, however, just Searle's principle of expressibility that hinting makes difficult to accept. Another well-known aspect of his philosophy of language runs the same risk. The fact that hinting is a clear case of language use with regard to

---

2 For interesting arguments against the principle of expressibility which do not rely on considerations regarding hinting, see Binkley (1979) and Gazdar (1981).

which a discrepancy must be recognised to exist between what a sentence intrinsically means and what a speaker means by uttering that sentence is of some help in assessing claims to the effect that two expressions have the same intrinsic meaning. In particular, it justifies one in saying that two expressions claimed to have the same intrinsic meaning do not, in fact, have the same intrinsic meaning if the one can, while the other can never, be used in order to give a particular hint (since, if their intrinsic meanings were the same, a speaker should either be able to use both or be unable to use either as a means of conveying something in excess of what he says). A claim that can be fruitfully assessed in this light is the one implicit in Searle's theory of proper names[3] (indeed, in the family of theories of proper names to which Searle's belongs).

Searle's theory, to briefly summarise it, is that the meaning of each proper name is the same as the meaning of a cluster of identifying descriptions of the individual to which it refers. Thus, the word "Aristotle" has a meaning (just as the word "helicopter" does), and its meaning is the same as the meaning of a cluster of identifying descriptions of Aristotle (for example, "The inventor of syllogistic", "The author of *Nicomachean Ethics*", "The teacher of Alexander the Great", and so on).

Searle's theory (along will all other desciptivist theories of names) has been famously attacked by Kripke (1972, 1980), but the above-mentioned considerations regarding hinting provide the opportunity for an additional line of attack. Consider the following situation: a person *B* finds another person *A* reading a book and asks *A* who the author of the book is. *A*, who is in a playful mood, does not reveal who the author is (*we* know it is Aristotle), but asks *B* to guess. *B* agrees to play the game, but asks for some hints. It is clear that if, in this situation, *A* was to say,

(11)   It's Aristotle,

he would *not* be described as having given a hint. It is equally clear, on the other hand, that if *A* was to say,

(12)   It's the inventor of syllogistic.

and, a little later,

---

3 Searle (1958; 1969: 162–174). See also Searle (1983: 231–261).

(13)   It's the author of *Nicomachean Ethics*.

and, still later,

(14)   It's the teacher of Alexander the Great,

and so on until he would have gone through the totality of identifying descriptions of Aristotle, he *would* be described as having given a (quite generous) hint. But this situation ought to be impossible if Searle's theory of proper names was right. For, according to Searle's theory, the sentence using Aristotle's name *has the same meaning* as the inclusive disjunction of sentences where Aristotle's name is replaced by whatever appropriate identifying descriptions of Aristotle there are. Therefore, either the sentence using Aristotle's name should constitute just as good a *hint* as the inclusive disjunction of those sentences, or the inclusive disjunction of those sentences should constitute just as good a *non*-hint as the sentence using Aristotle's name. Since neither of these possibilities is actualised, it seems that something about Searle's account of proper names has gone badly wrong.

Chapter 16
How to test a test for perlocutionary act names

Steven Davis (1979) assumes—following Austin (1975)—that a perlocutionary act is one whereby, as a result of performing an illocutionary act, we succeed in bringing about "certain effects on the thoughts, actions, or feelings of our hearers" (Davis 1979: 242). Davis is aware of the fact that the grammatical tests Austin once tried to apply for distinguishing perlocutionary from non-perlocutionary act names fail; so, after reviewing the reasons why they do, he goes on to propose a complex test of his own which, he believes, does not. "I would like to propose", he writes, "a linguistic test which, I believe, does distinguish perlocutionary act verbs from other verbs" (Davis 1979: 237).

The test is defined by reference to the following frames, where S is "a variable for a designation for a speaker", H is "a variable for a designation for a hearer", φ "ranges over illocutionary and propositional act verbs", and X and Y can be substituted by "any linguistic expressions or nothing as long as grammaticality is preserved" (Davis 1979: 237):

(i)   By φ-ing X S ψ-s H Y.
(ii) S's φ-ing X ψ-s H Y.
(iii) H was ψ-ed Y by S's φ-ing X.

The test runs as follows: "A verb substituted for ψ is a perlocutionary act verb just in case there are substitution instances for the other variables which render [(i), (ii) and (iii)] grammatical" (Davis 1979: 237).

In order to show that verbs strongly suspected to be perlocutionary, such as "to amuse", do qualify as perlocutionary by his test, Davis points out that all of the following sentences are grammatical:

(1)   By objecting to her criticism Abel amused Mabel.
(2)   Abel's objecting to her criticism amused Mabel.
(3)   Mabel was amused by Abel's objecting to her criticism.

And in order to show that verbs strongly suspected not be perlocutionary, such as "to suggest", do not qualify as perlocutionary by his test, he points out that not all of the following sentences are grammatical:

(4)   By telling Mabel it was late Abel suggested to her that she leave.
(5)   Abel's telling Mabel it was late suggested to Mabel that she leave.
(6) * Mabel was suggested that she leave by Abel's telling her it was late.

These examples create a problem that must be faced before any attempt at evaluating Davis' test is undertaken. The problem stems from the fact that they cannot possibly show what Davis claims they show, since the first two of the sentences cited for both "to amuse" and "to suggest" are not substitutions instances of the frames they are said to be substitutions instances of: in these frames, ψ is coded to past tense rather than to present tense morphemes. Since it is not evident whether Davis would opt for the past tense course or for the present tense course in order to remedy this particular anomaly, it is not evident which of the two corresponding versions of the test should be accepted as authoritative. Consequently, if one wants to show beyond doubt the test's incapacity to accomplish the task it was expected to accomplish, one must show this by reference to each one of the versions in question. And the most economical way of doing this is by means of counterexamples whose grammaticality or ungrammaticality would remain unaffected by changes from past to present or from present to past in the relevant substitutes for ψ. We are now in a position to evaluate Davis' test for distinguishing perlocutionary act verbs from other verbs.

The test fails to establish this distinction because, in each of its two possible versions, (A) it marks as perlocutionary act verbs certain verbs which are not action verbs at all, (B) it marks as perlocutionary act verbs two types of action verbs that are clearly not perlocutionary act verbs, and (C) it marks as non-perlocutionary act verbs certain action verbs that are clearly perlocutionary act verbs. Let us establish each of these points in order.

(A) Leaving a person entirely unaffected in its thoughts, feelings, and actions is certainly not *doing* something (and it is even less bringing about certain effects on that person's thoughts, feelings, and actions). Davis, however, must admit that leaving a person entirely unaffected in its thoughts, feelings, and actions *is* doing something (and, what is more, that it *is* bringing about certain effects on that person's thoughts, feelings, and actions). This happens because the following sentences, which are substitution instances of (i), (ii) and (iii) respectively (for either past or present tense versions of ψ in the first two) are all grammatical, and

because verbs grammatically substituted for ψ in (i), (ii) and (iii) are, according to the test, perlocutionary act verbs:

(7) By pointing out that there are still chances for world peace Abel left/leaves Mabel entirely unaffected in her thoughts, feelings, and actions.

(8) Abel's pointing out that there are still chances for world peace left/leaves Mabel entirely unaffected in her thoughts, feelings, and actions.

(9) Mabel was left entirely unaffected in her thoughts, feelings, and actions by Abel's pointing out that there are still chances for world peace.

Since, therefore, the test makes it necessary to accept as perlocutionary act verbs certain verbs which are not even action verbs, it cannot itself by accepted.

(B) Let us now turn to *action* verbs which are not *perlocutionary* act verbs but which qualify as *perlocutionary* act verbs by the test under examination. Speech act verbs of this sort can be either illocutionary or non-illocutionary, as the following cases illustrate.

"To give permission" is a clear case of an illocutionary act verb and, correspondingly, an equally clear case of a non-perlocutionary act verb (it is perfectly consistent, in fact, to say that one gave someone permission to do a certain thing while denying that one has thereby had any effect on that person's thoughts, feelings, or actions). This verb, however, qualifies as a *perlocutionary* act verb by Davis' test, since it can be grammatically substituted for ψ in (i), (ii) and (iii)—with either past or present tense forms in (i) and (ii)—as the following sentences show:

(10) By declaring to her "You may leave" Abel gave/gives Mabel permission to leave.

(11) Abel's declaring to her "You may leave" gave/gives Mabel permission to leave.

(12) Mabel was given permission to leave by Abel's declaring to her "You may leave".

"To give permission", then, represents a first variety of non-perlocutionary act verbs which qualify as perlocutionary act verbs by Davis' test, and which thereby disqualify the test in question.

"To echo" is clearly not an illocutionary act verb and it is clearly not a perlocutionary act verb either (it is perfectly consistent, in fact, to say that one echoed someone while denying that one has thereby had any effect on that person's

thoughts, feelings, or actions). This verb too, however, qualifies as a perlocutionary act verb by Davis' test, since it can be grammatically substituted for ψ in (i), (ii) and (iii)—with either past or present tense forms in (i) and (ii)—as the following sentences show:

(13)   By arguing that people like Pavel are incorrigible Abel echoed/echoes Mabel.

(14)   Abel's arguing that people like Pavel are incorrigible echoed/echoes Mabel.

(15)   Mabel was echoed by Abel's arguing that people like Pavel are incorrigible.

"To echo", then, represents a second variety of non-perlocutionary act verbs which qualify as perlocutionary act verbs by Davis' test, and which thereby disqualify the test in question.

(C) Let us finally turn to action verbs which *are* perlocutionary act verbs, but which qualify as *non*-perlocutionary act verbs by the test under examination.

As Davis observes in the course of discussing an independent topic (Davis 1979: 238), frightening someone (by saying something) and causing someone to become frightened (by saying something) are both perlocutionary acts (it is, in fact, hardly consistent to say that one frightened someone, or that one caused someone to become frightened, and deny that one has thereby had any effects on that person's thoughts, feelings, or actions). What Davis has failed to observe, however, is that, although "frighten x"' can be grammatically substituted for ψ throughout (i), (ii) and (iii), "cause x to become frightened" *cannot* be grammatically substituted for ψ throughout (i), (ii) and (iii). The following examples illustrate this last point.

(16)   By telling her that the Vikings are coming Abel caused/causes Mabel to become frightened.

(17)   Abel's telling her that the Vikings are coming caused/causes Mabel to become frightened.

(18) * Mabel was caused to become frightened by Abel's telling her that the Vikings are coming.

Since "cause x to become frightened" cannot be grammatically substituted for ψ throughout (i), (ii) and (iii), Davis' test cannot accept it as a perlocutionary act verb. And since, by Davis' as well as by everyone else's admission, it is as much

a perlocutionary act verb as "frighten *x*" is, it is the test itself that cannot be accepted.

In short, Davis' test for distinguishing perlocutionary act verbs from other verbs fails in each version and in every respect, and the problem of giving grammatical guarantees for this distinction remains accordingly where it always was, namely, in the list of problems awaiting solution.

Would the search for such a solution be a reasonable enterprise? I very much doubt that it would. If the *semantic* contrast between illocutionary and perlocutionary verbs—namely, that the former do not whereas the latter do denote instigations of changes in the thoughts, feelings, or actions of addressees—can effectively separate the two verb classes, the question whether these two *semantically* separable classes have or do not have, as classes, proprietary grammatical properties has no obvious theoretical motivation or interest. And there is, to my knowledge, no compelling evidence that the semantic contrast *cannot* be relied upon to separate the two classes. Indeed, it may well be that the cases where a verb can *appear* difficult to categorise as semantically illocutionary or semantically perlocutionary are in fact cases of ambiguity (specifically, of *polysemy*), in which the verb has distinct illocutionary and perlocutionary senses. To give just two examples, one may, at first glance, find it difficult to tell whether the verb "to insult" is semantically illocutionary or perlocutionary; however, by the time one notices that a sentence like (19) is *not* contradictory,

(19)   I insulted him repeatedly, but he wouldn't be insulted.

one is forced to acknowledge that the verb "to insult" has two distinct *senses*, one of which (the sense in which it denotes the actual *instigation* of a certain type of emotional effect on a person) is clearly perlocutionary, and the other one of which (the sense in which it denotes the *attempt to instigate* that type of emotional effect on a person) is clearly illocutionary. Similarly, it may, at first glance, appear difficult to decide whether the verb "to encourage" is semantically illocutionary or perlocutionary; however, by the time one notices that a sentence like (20) is *not* contradictory,

(20)   He encouraged her many times to submit a proposal, but she was not encouraged.

one is forced to acknowledge that the verb "to encourage" has two distinct *senses*, one of which (the sense in which it denotes the actual *instigation* of a certain type of cognitive or behavioural effect on a person) is clearly perlocutionary, and the

other one of which (the sense in which it denotes the *attempt to instigate* that type of cognitive or behavioural effect on a person) is clearly illocutionary. It is in the detailed description of such cases of regular polysemy, and in the exploration of their causes and consequences, that, I believe, the real theoretical interest lies.

Chapter 17
Speaker meaning, sentence meaning, and metaphor

1 Introduction

Two widely held assumptions in contemporary discussions of meaning are, first, that a distinction deserves to be drawn between what sentences of natural languages mean and what speakers of those languages mean by uttering those sentences; and, second, that this distinction largely determines the distinction between semantics and pragmatics as domains of inquiry, with semantics being dedicated to the analysis of sentence meaning and pragmatics being concerned with the analysis (usually along lines inspired by the Gricean theory of conversational implicatures) of the the various kinds of meaning that are instances of speaker meaning rather than of sentence meaning.

Metaphorical meaning is among the kinds of meaning that those who subscribe to these assumptions (including Grice himself) regard as a kind of meaning that is an instance of speaker meaning rather than of sentence meaning. But since metaphorical meaning can hardly be taken to be a pretheoretically *obvious* case of meaning that is speaker-based rather than sentence-based, and since there is a considerable number of not obviously flawed theories, both ancient and modern, where the phenomenon of metaphor has been regarded as a semantic rather than as a pragmatic phenomenon,[1] attempts to analyse metaphor within the context of the two assumptions can—and must—be evaluated in at least two ways. First, by considering how well they manage to distinguish metaphor from *other* presumed cases of speaker meaning, assuming that it *is* a case of speaker meaning. And second, by considering how well they motivate their initial assumption that metaphor *is* a case of speaker meaning rather than of sentence meaning—in other words, that it is a pragmatic rather than a semantic phenomenon. In what follows, I will be referring to these two questions as the "internal" and the "external" question, respectively.

---

1 Cf., among relatively recent works, Cohen and Margalit (1972), Cohen (1979), Lakoff and Johnson (1980), Ross (1981), MacCormack (1985), Kittay (1987).

My purpose in this essay is to examine, from the two perspectives just indicated, some claims of John Searle's paper on metaphor (Searle 1979: 76–116), which is one of the best-known attempts to analyse metaphorical meaning as a special case of speaker meaning rather than of sentence meaning, within what Searle regards as a broadly Gricean framework. I will argue that Searle's attempt—which, in the respects that are relevant here, is representative of most other pragmatic approaches to metaphor[2]—fails both as an attempt to distinguish metaphor from other presumed cases of speaker meaning, and as an attempt to motivate the assumption that metaphor *is* in fact a case of speaker meaning. To the extent that the essay succeeds, then, it suggests that, as far as the linguistically central phenomenon of metaphor is concerned, the widespread tendency to reinterpret as many semantic phenomena as possible in pragmatic terms, may have to be firmly resisted.

2 The internal question

A speaker speaks *literally*, according to Searle, just in case what *he* means by uttering a sentence is identical with what the sentence he utters *itself* means. (Thus, a speaker who, in saying, "There are prime numbers", means that there are prime numbers is a speaker who speaks literally, since what *he* means is identical with what the sentence he utters—namely, the sentence "There are prime numbers"— itself means.) Cases where what a speaker means by uttering a sentence is not identical with what the sentence he utters itself means come, according to Searle, in two varieties. On the one hand, a speaker who, in uttering a sentence, means *not only* what the sentence he utters itself means but, *in addition*, something else as well is a speaker who speaks *indirectly*. (For example, a speaker who, in saying to his hearer, "Can you tell me the time?", means both that he would like to know whether the hearer can tell him the time *and* that he would like to know what the time is, is a speaker who speaks indirectly, since, although he does mean what the sentence he utters itself means—namely, that he would like to know whether the hearer has the ability to tell him the time—means, in addition, something else that is not meant by the sentence he utters—namely, that he would like to know what the time actually is.) On the other hand, a speaker who, in uttering a sen-

---

2 And also of some well-known approaches which do not label themselves "pragmatic", such as Davidson's (1978).

tence, does not mean *at all* what the sentence he utters itself means but something different altogether is a speaker who speaks *figuratively*. (For example, a speaker who, in saying, "Sally is a block of ice", means that Sally is unemotional, or a speaker who, in saying, "Sally was very kind to me", means that Sally was very rude to him, are speakers who speak figuratively, since they do not mean *at all* what the sentences they utter themselves mean—namely, that Sally is a block of ice or that Sally was very kind, respectively—but something else altogether—namely, that Sally is unemotional and that Sally was very rude, respectively.) Finally, a speaker who speaks figuratively—that is to say, a speaker who does not mean at all what the sentence he utters means, but something different altogether—can, according to Searle, be doing the one or the other of two different kinds of things. He may be speaking *ironically*, in which case not only does he not mean at all what the sentence he utters means but means the *opposite* of what the sentence he utters means; or he may be speaking *metaphorically*, in which case he does not mean at all what the sentence he utters means, but neither does he mean the *opposite* of what the sentence he utters means. (Thus, a speaker who, in saying, "You have been very kind to me, Sally", means that Sally has been very rude to him is a speaker who speaks ironically, because not only does he not mean what the sentence he utters means, but means the opposite of what that sentence means—given that a person's being rude is the opposite of that person's being kind; whereas a speaker who, in saying "Sally is a block of ice", means that Sally is unemotional, is a speaker who speaks metaphorically, because, though he does not mean at all what the sentence he utters means, neither does he mean the opposite of what that sentence means—given that an object's being a block of ice is not the opposite of that object's being unemotional.)

For the sake of argument, let us now assume, with Searle, that metaphorical meaning is strictly a matter of speaker meaning rather than of sentence meaning. The internal question to ask about Searle's proposal can then be split into two sub-questions. First, whether it provides a basis for a proper distinction between metaphor and irony. And second, whether it provides a basis for a proper distinction between metaphor and irony, on the one hand, and non-literal but not figurative (that is, indirect) uses of sentences, on the other.

If Searle's way of distinguishing ironical utterances from metaphorical utterances were correct, there would not be any utterances that one could properly describe as *simultaneously* ironical and metaphorical, since saying of an utterance that it is simultaneously ironical and metaphorical is, within Searle's theory, equivalent to saying that what the speaker of that utterance means both *is* (in view of its ironical character) and *is not* (in view of its metaphorical character)

opposite to what the sentence he utters itself means. However, there certainly exist utterances that are properly describable as being ironical and metaphorical at the same time. If, for example, I wish to suggest that a certain person that someone has just described as very dependable is not dependable at all, I may succeed in doing so not only by saying, ironically, "He is very dependable indeed!", but also by saying, equally ironically,

(1)   Sure, he is a rock!

But this last utterance would not only be ironical but also metaphorical—as would be any other utterance in which an animate object would be described as an inanimate one. And since, on Searle's theory, speaking ironically *is* meaning the opposite of what the sentence one utters means whereas speaking metaphorically is *not* meaning the opposite of what the sentence one utters means, anyone uttering the metaphorical sentence in (1) ironically would have to be counted as meaning and not meaning at the same time the opposite of what (1) means. Similarly, if I wish to suggest that a man that someone has just described as very brave is, in fact, not brave at all, I may succeed in doing so not only by saying, ironically, "That man is very brave indeed!", but also by saying, equally ironically,

(2)   That man is a lion, of course!

But this last utterance would be not only ironical but also metaphorical, as would be any other utterance in which a human being would be presented as belonging to a class of non-human beings. In these and in numerous other cases of the same kind, then, Searle's account would be forced to translate the obviously correct observation that the same utterance can be simultaneously an instance of metaphor and an instance of irony into the incoherent claim that what the speaker of that utterance means both is and is not opposite to what the sentence he utters means. Consequently, Searle's proposed way of distinguishing ironies from metaphors cannot be maintained, even if we grant that the latter are, like the former, manifestations of speaker meaning rather than of sentence meaning (an assumption that the examples just considered already make suspect, since they suggest that metaphorical interpretations are context-free in a way in which ironical interpretations are not: one needs quite specific information about the context in order to determine whether "He is a rock!" or "That man is a lion!" are meant ironically or non-ironically; but no such information is required in order to deter-

mine that they are meant *metaphorically*, and it is, indeed, hard to think of a context in which the metaphors they contain would have *not* been identified and in which they would *still* be counted as acceptable utterances).

Let us now turn to Searle's proposed elucidation of the difference between figurativity and indirection. On his account, what distinguishes an utterer who speaks figuratively from an utterer who speaks indirectly is that, although they both mean certain things that are different from those meant by the sentences they utter, the former does not mean *at all* the thing meant by the sentence he utters, whereas the latter means *in addition* the thing meant by the sentence he utters. If Searle's way of drawing this distinction was correct, then, there would not exist any utterers that one could properly describe as simultaneously speaking figuratively and indirectly, since that description would entail that they simultaneously mean and do not mean what the sentences they utter mean. However, there certainly exist utterers whose utterances can properly be described as being simultaneously instances of figurativity and instances of indirection. For example, just as I can indirectly suggest to someone that her baby needs a bath by saying to her "Your baby is full of dirt", I can indirectly suggest to her that her baby needs a bath by saying to her,

(3)   Your baby has become a piglet.

But the fact that this last utterance can be the vehicle of an indirect suggestion hardly prevents it from being a vehicle of metaphor, and it is, in fact, only when its intended metaphorical interpretation has been established that its indirect force of suggestion can be properly attributed. Similarly, just as I can indirectly express my admiration for a certain painting by asking the question, "Isn't that painting admirable?", I can indirectly express my admiration for the same painting by asking the question,

(4)   Isn't that painting a jewel?

But the fact that the latter utterance is, just like the former, a possible vehicle of indirection hardly prevents it from being, unlike the former, a vehicle of metaphor, and it is, once again, only when its intended metaphorical interpretation has been determined that its indirect force can be properly attributed. Faced with these and numerous other cases of the same sort, however, Searle would be forced to translate the obviously correct observation that a certain utterance is at the same time an instance of figurativity and an instance of indirection into the incoherent claim that its speaker has achieved the impossible task of having

meant and of not having meant at the same time what was meant by the sentence he has uttered—for, the first of these features is, on Searle's account, a necessary feature of every instance of indirection and the second a necessary feature of every instance of figurativity. It seems, then, that Searle's proposed way of distinguishing between figurativity and indirection is no more successful than his proposed way of distinguishing between the two main kinds of figurativity that he takes to be represented by irony and by metaphor, respectively. Consequently, the two internal sub-questions that his pragmatic account of metaphor was supposed to be able to resolve can certainly not be regarded as having been successfully resolved. Let us now proceed to the external question that any such conception would have to be able to answer in the affirmative if it were to ensure that it is not fundamentally misconceived: is there any good reason for thinking that metaphor *is* a matter of speaker meaning rather than of sentence meaning—in other words, that it *is* a pragmatic rather than a semantic phenomenon?

3 The external question

Searle's answer to that question appears to be that it is *obvious* that metaphorical meanings are speaker-based rather than sentence-based. But the fact that the question whether a given conveyed meaning is speaker-based rather than sentence-based has obvious answers in certain cases does not entail that it has an obvious answer in every possible case, and this, as noted, is especially clear in the case of metaphor, where many people before and after Searle have regarded as far from obvious the thing that he takes to be obvious. It would appear, then, that reference to some independently available criterion would be needed to settle this fundamental question. And, fortunately for Searle, the Gricean framework within which he sees himself as operating provides a criterion, the so-called "cancellability test", which does purport to properly motivate decisions as to whether a given conveyed meaning is speaker-based or sentence-based. It appears, however, that Searle has neglected to apply the cancellability test before making his decisions concerning the analysis of metaphor. For, as I will now argue, that test, applied to relevant cases, clearly contradicts his (and many others', including Grice's own) assumption that metaphorical meanings are speaker-based rather than sentence-based.

The principle of the cancellability test—whose basic idea comes from Grice (1975, 1989), and which was employed by Searle himself on other occasions (for example in Searle 1979: 30–57)—is simple and, I think, sound: If a speaker who has uttered a sentence *S*, and has been interpreted as having thereby meant that

p, can without linguistic oddity *cancel* that interpretation of his utterance, then his having meant that *p* by *S* is not part of what *S* itself means; if, on the other hand, a speaker who has uttered a sentence *S*, and has been interpreted as having thereby meant that *p*, cannot without linguistic oddity *cancel* that interpretation of his utterance, then his having meant that *p* by *S* is part of what *S* itself means. (Thus, if a speaker who has said, "I am a man", has been interpreted as having meant, among other things, that he is human, his having meant that he is human must be taken to be part of what the sentence he has uttered *itself* means, since he could *not* without oddity have said, "I am a man, but that doesn't mean that I am human"; on the other hand, if a speaker who has said, "I am a man", has been interpreted as having meant, among other things, that he would like to meet a woman, his having meant that he would like to meet a woman cannot be taken to be part of what the sentence he has uttered *itself* means, since he *could* without oddity have said, "I am a man, but that doesn't mean I would like to meet a woman".)

Now, many of the *non*-metaphorical cases where, for Searle as for many others, a speaker's meaning diverges from a sentence's meaning are easily confirmed to be cases of such divergence by the cancellability test. Thus, the test shows that, if a speaker who says, "I am hungry", is interpreted as asking for food, the interpretation in question cannot legitimately be supposed to be part of what the sentence he uttered *itself* means, since the speaker could without oddity have cancelled that interpretation by producing an utterance like (5):

(5)   I am hungry—but please don't give me any food: I am on a diet.

Similarly, the test shows that, if a speaker who says, "Why should I ever divorce my wife?", is interpreted as expressing the opinion that there are no reasons why he should divorce his wife, the interpretation in question cannot legitimately be supposed to be part of what the sentence he uttered *itself* means, since the speaker could without oddity have cancelled that interpretation by producing an utterance like (6):

(6)   Why should I ever divorce my wife?—I don't mean to suggest that no reasons could ever be found; I simply want to be told what these reasons are.

What Searle and many others (including Grice) have failed to notice, however, is that, when applied to metaphorical utterances, the cancellability test gives re-

sults that contradict the thesis that metaphorical meanings are speaker-based rather than sentence-based. Suppose, to adapt one of Searle's favourite examples, that a speaker says, "My wife is a block of ice", and is interpreted as meaning metaphorically that his wife is, say, unemotional. If this or any other metaphorical interpretation is *not* part of what the sentence the speaker utters means, then the speaker should be able to cancel without oddity all interpretations of his utterance except the one that is strictly identical to its literal meaning. Suppose, then, that the speaker attempts to block all metaphorical interpretations of his utterance, by speaking as if his wife is literally a block of ice and nothing else—by saying, for example,

(7)   My wife is a block of ice—so, please help yourselves to my wife if you need ice for your drinks.

This utterance, I submit, would elicit the one or the other of two types of reaction. Either it would be immediately rejected as semantically anomalous, or it would be accepted as semantically well-formed *provided that* one would have managed to interpret the speaker as *still* speaking metaphorically when describing his wife as something that his interlocutors might choose to add to their drinks. But this means that the attempt to cancel without oddity every metaphorical interpretation of "My wife is a block of ice" cannot possibly succeed: either the result will be an utterance that is rejected as semantically odd, or it will be an utterance that is accepted as semantically not odd precisely because the metaphor has *not* been cancelled. And this in turn means that, as far as the cancellability test is concerned, metaphorical interpretations of uttered sentences must be supposed to be just functions of what the sentences *themselves* mean rather than functions of what speakers may choose to mean by uttering those sentences. Suppose, to take one more, familiar, example, that the sentence "Time is money" receives, on a particular occasion of utterance, one of its usual metaphorical interpretations. If these interpretations are not determined by what *it* means but rather by what its speaker has chosen to mean by uttering it, then its speaker should be in a position to block without oddity all metaphorical interpretations of his utterance by going on to speak as if time was, literally, a kind of money and nothing else—by saying, for example,

(8)   Time is money—so, how much of your time have you got in your bank account?

But this utterance would be either rejected as semantically anomalous or accepted as semantically well-formed *provided that* its hearer would have managed to interpret its speaker as *still* speaking metaphorically when describing time as something that can be deposited in a bank account. This means that the result of the attempt to prevent without oddity the metaphorical interpretation of "Time is money" will be either an utterance that *is* odd or an utterance whose metaphorical interpretation has *not* been prevented. And since, according to the cancellability test, it is sentence meanings, rather than speaker meanings, that cannot be prevented without oddity, the test's verdict must, as in the previous case, be that the metaphorical meaning of "Time is money" resides in what *it* means rather than in what any speaker might have chosen to mean by uttering it.

Since the view that metaphor is a matter of speaker meaning rather than of sentence meaning can hardly be regarded as obviously correct, since the cancellability test has been devised precisely in order to help deciding unobvious cases of this kind, and since it is a test that appears both to rest on sound assumptions and to give the expected results in cases where the distinction between speaker meaning and sentence meaning *is* obvious, the only reasonable interpretation of the above results is the interpretation according to which they do in fact show what they appear to show—namely, that metaphors can legitimately be regarded as functions of what sentences *themselves* mean, rather than as functions of what speakers choose to mean by uttering them. (And if this is so, of course, Searle's previously encountered difficulties in distinguishing metaphorical meaning from what he regards as *other* types of cases of *speaker* meaning should hardly appear surprising: for if these other cases *are* cases of speaker meaning whereas metaphors are *not*, it is no wonder that attempts to distinguish the former from the latter as different species of the same genus cannot succeed.)

4 Conclusion

My purpose in this essay has been twofold. On the one hand, I have tried to show that, even if we assume that metaphorical meanings are functions of what speakers, as opposed to sentences, mean, Searle's proposed bases for distinguishing metaphorical meanings from other kinds of speaker meanings are unreliable, since they lead to contradictory statements both in those cases in which an utterance can be simultaneously an instance of metaphor and an instance of irony and in those cases in which an utterance can be simultaneously an instance of figurativity and an instance of indirection. On the other hand, I have tried to show that Searle's assumption that metaphors *are* functions of what speakers, as opposed

to sentences, mean, is incorrect, at least when viewed in the light of the test that is most widely accepted as affording a reasoned decision on the question whether a given conveyed meaning is or is not a function of what a sentence, as opposed to a speaker, means. If my arguments are well taken, then, they will be of interest both to those who have been led to suppose that the distinction between speaker meaning and sentence meaning holds the key to the analysis of metaphor, and (for different reasons) to those who have long suspected that metaphor is too fundamental a feature of natural languages to admit of a simple pragmatic, as opposed to a complex semantic, treatment. [3]

---

3 For important criticisms of aspects of Searle's account of metaphor other than those examined in this essay, the reader will profitably consult Cohen (1979) and Cooper (1986). An analysis of metaphor (and of related topics) that is directly inspired by Searle, and to which therefore the arguments presented here directly apply, is Vanderveken (1991). These arguments also apply to the pragmatic analysis of metaphor proposed in Fogelin (1989), even though that analysis is considerably more sophisticated than—and at certain points justly critical of—Searle's own.

Chapter 18
Voices and noises in the theory of speech acts

1 Introduction

The present essay reviews two recent collective volumes on speech acts and related topics: *Speech Acts, Mind and Social Reality: Discussions with John R. Searle*, edited by Günther Grewendorf and Georg Meggle (Grewendorf and Meggle 2002), hereafter abbreviated as G&M; and *Essays in Speech Act Theory*, edited by Daniel Vanderveken and Susumu Kubo (Vanderveken and Kubo 2002), hereafter abbreviated as V&K. Section 2 of the essay is devoted to G&M, while Section 3 is devoted to V&K.

Though different in important respects, the two volumes have a common feature that imposes a clear restriction on the scope of the present review, namely, that (contrary perhaps to what some of their readers might be expecting or wishing) neither of them aspires to offer a comprehensive view of different approaches to fundamental issues of speech act theory (of the sort that was attempted in Tsohatzidis 1994 a decade ago) and neither of them endeavours to establish systematic contact with innovative contemporary research in the theory of meaning that is either directly couched in speech act theoretic terms (see, for example, the approach developed in Alston 2000) or indirectly draws on speech act theoretic notions (see, for example, the approach summarised and expanded upon in Brandom 2000). This fact does not make either volume an appropriate occasion for in depth discussion of general issues of speech act theory or of its place within the theory of meaning, and the remarks to follow will accordingly have to be more narrowly focused.

The main difference between the volumes, as far as the present appraisal of them is concerned, is that the first has a unifying, if complex, theme, is moderately well edited, and contains a not insignificant number of valuable papers, whereas the second does not appear to have a unifying theme, is very inadequately edited, and contains a large number of papers whose significance is at best unobvious and at worst non-existent. In the separate presentations that follow, the emphasis will be on evidence supporting the above comparative characterisation, though incidental suggestions concerning the treatment of some of the

issues raised by individual papers will also be offered. Unless otherwise indicated, parenthetical page references in Section 2 are to G&M, and parenthetical page references in Section 3 are to V&K.

2 Remarks on G&M

The G&M volume assembles papers descending from a 1999 Bielefeld colloquium on the philosophy of John Searle, and, as its title indicates, it includes essays in all three of the main research areas to which Searle has made significant contributions: philosophy of language, philosophy of mind, and social ontology. The book's corresponding three core parts (entitled, respectively, 'Speech Acts', 'Mind' and 'Social Reality') contain eighteen original essays, of which eight are allocated to the philosophy of language part, five to the philosophy of mind part, and five to the social ontology part.

The three core parts are preceded by a part called 'Introduction', and followed by a part called 'New perspectives'. The 'Introduction' part consists of a paper by Searle ("Speech acts, mind and social reality") in which he gives a very brief outline of his published work in the three research areas to which the volume is dedicated. (This paper is followed by the transcript of a short conversation between Searle and Ralf Stoecker, in which Searle repeats some of the points made in the paper, and briefly clarifies some others). The 'New perspectives' part consists of a second paper by Searle ("The classical model of rationality and its weaknesses"), which provides a glimpse of his recent (and, at the time of the colloquium, unpublished) work in the latest of his areas of philosophical interest, the theory of rational action.

Although the editors' idea of asking Searle to contribute an opening and a closing paper to a volume of essays devoted to his philosophy was, in itself, obviously appropriate, the degree to which these particular opening and closing papers actually enhance the volume's coherence or originality is minimal. The opening paper will probably be unnecessary for readers already acquainted with Searle's major works, and it will certainly be insufficient for readers wholly unacquainted with those works. Perhaps, then, its intended beneficiaries are primarily readers who happen to be familiar with only one of the three areas of Searle's philosophical interest and who want to know both what his distinctive positions in the remaining two areas are and how his central claims in each one of the three areas impose constraints on the evaluation of his contributions to the other two. The opening paper, however, is far too brief to allow such readers to form an informed opinion on these matters, and Searle is obviously justified

when he disarmingly declares, toward the end of the paper, "I do not wish any critic of my views to suppose that this brief summary (...) could be a target for a critical assessment of my work" (p. 16). In any case, Searle has already published a short introductory book, *Mind, Language and Society* (Searle 1998), whose purpose is precisely to provide interconnected and detailed summaries of his contributions to the philosophy of language, to the philosophy of mind and to social ontology, and that book (whose study might be supplemented with the study of the Searlian interviews published in Faigenbaum 2001) could serve the needs of wholly or partly unfamiliar readers much more usefully than the volume's opening paper (which actually reads, for the most part, like an extended table of contents of Searle's introductory book). As for the volume's concluding paper, it might have been interesting, at the time of the colloquium, to hear Searle launching some of his then unpublished theses on the idea of rationality, but these theses, together with the theoretical framework within which their significance could be best understood and assessed, have in the meantime become available in his book *Rationality in Action* (Searle 2001), published one year before the G&M volume, and it is obviously to that earlier book rather than to the volume's concluding paper (which simply reproduces part of the earlier book's first chapter) that interested readers should direct themselves.

Searle's opening and closing papers contain no trace of a response to the various queries, comments and criticisms addressed to him by the authors of the eighteen papers that form the main body of the volume. This is disappointing, and does not help the volume compare favourably with the only previous collective work of similar scope, *John Searle and his Critics* (Lepore and Van Gulick 1991), which not only contains some high-powered critical essays on various aspects of Searle's philosophy, but also includes Searle's detailed replies to each one of his critics. The editors cannot, of course, be held responsible for Searle's unwillingness to respond in print to any of the colloquium papers; it would appear to belong to their responsibilities, however, to ensure two things: first, that prospective readers do not acquire the false impression that the volume does include such responses; and second, that prospective readers receive some introductory information as to what the content of each one of the eighteen papers is, and how it purports to advance our understanding of its chosen topic. Unfortunately, neither of these responsibilities has been assumed. On the one hand, the volume is subtitled *Discussions with John R. Searle*, even though it contains absolutely no instance of Searle discussing with any of the contributors any of the points that they raise. On the other hand, the editors do not supply any kind of introduction outlining what each of the eighteen papers is attempting to do, or how their individual aims are connected both to each other and to Searle's work,

and they do not even equip the volume with an index that some readers might wish to use in trying to trace such connections for themselves. Perhaps, then, the only credit that the editors would expect to be given, as far the presentation of the volume is concerned, is that they have ensured, first, that it is generally free of typographical errors and, second, that the language of those of its contributors who are not native speakers of English, is, though not always idiomatic, almost always comprehensible.

The quality of the papers in the volume's three core parts is variable, though those that are worth at least one reading constitute the majority. I will first go briefly through the papers in the 'Mind' and 'Social Reality' parts, which together take up roughly half of the volume's space, and then turn to the papers in the 'Speech Acts' part, which occupy the other half.

The 'Mind' part consists of an uncharacteristically unfocused paper by Avrum Stroll, a useful essay by Wolfgang Lenzen, a solid contribution by Martine Nida-Rümelin, a largely irrelevant essay by Thomas Bartelborth and Oliver Scholz, and a minimally relevant but maximally presumptuous paper by Thomas Roeper. The purpose of Stroll's paper ("Identification and misidentification") is to examine whether Searle has succeeded, in *Intentionality* (Searle 1983), to defend his internalist account of mental and linguistic content against the externalist challenges arising from the so-called 'new theory of reference', especially as the latter has been developed in Hilary Putman's early work on the semantics of proper names and natural kind terms (Putnam 1975). Stroll claims, in effect, (a) that Searle has successfully defused the externalist challenges as far as proper names are concerned, (b) that he cannot defuse the externalist challenges as far as certain natural kind terms are concerned, and (c) that Putnam's own externalist account of certain *other* natural kind terms is itself problematic. Claim (a) is not, given the present state of the debate on proper names, developed satisfactorily: Stroll simply declares himself convinced that Searle has successfully defended his internalist account of proper names against externalist challenges, but he fails to discuss, or even to mention, the very detailed, and thoroughly negative, externalist examination of the Searlian arguments by Devitt (1990), and makes no attempt to produce any *new* arguments in favour of Searle's position that might be considered capable of neutralising Devitt's critique. Besides, the paper does not mention a paper by Stroll himself where Searle's account of proper names is explicitly claimed to rest on an assumption that, independently of one's position on the internalist/externalist dispute, is untenable (see Stroll 1998 and the reiteration of that claim in Chapter 8 of Stroll 2000), and this risks creating the inaccurate impression that Searle's account of proper names is actually ac-

ceptable to Stroll as it stands. Claim (b) is convincingly argued, though the assumption that there are people who would need to be convinced—that is, people who would believe that Searle has *any* sufficiently articulated account of natural kind terms, let alone one that could accommodate externalist challenges—appears to be unwarranted. Claim (c), which (judging from the amount of space devoted to it) seems closest to Stroll's interests, is, though quite important in itself, not exactly pertinent as stated, and thus potentially misleading: The full basis of that claim is to be found in previous writings of Stroll's that, again, he doesn't cite (see Stroll 1989 and the subsequent elaboration of that paper's ideas in Chapter 8 of Stroll 2000), in which he argues, on the one hand, that the 'scientistic' version of externalism associated with Putnam's early work provides an inadequate account of certain natural kind terms, and, on the other hand, that a non-'scientistic' version incorporating recognisably Wittgensteinian elements would supply a more adequate overall account of such terms. (For a general proposal as to how externalist and Wittgensteinian accounts of mental or linguistic content might be combined, see McCulloch 1995.) The Wittgensteinian direction that Stroll adopts in these writings may indeed be the most promising, but, as his present paper fails to make clear, it could hardly be a comforting direction *for Searle* to take, since the absolute *internalism* about content that Searle espouses is as far removed from early Putnamian 'scientistic' externalism as it is removed from late Wittgensteinian anti-'scientistic' externalism. It seems to me, then, that Stroll's contribution has been a missed opportunity to fully articulate one important family of problems that the Searlian account of content in *Intentionality* faces.

The papers by Wolfgang Lenzen and by Martine Nida-Rümelin examine aspects of those of Searle's contributions that are situated at the interface of the philosophy of mind and the philosophy of psychology. Lenzen's paper ("Intrinsic intentionality") is primarily a critical evaluation of some aspects of the so-called Chinese Room Argument that Searle presented in Searle (1980) and then in *Minds, Brains and Science* (Searle 1984) as part of his many-sided attack on computational theories of mind. Lenzen does not attempt to survey the extensive critical literature that the argument has triggered since its appearance, contenting himself to present his own reasons for, on the one hand, doubting that Searle's argument establishes its intended conclusion and, on the other hand, resisting the tendency of one of its prominent critics, Daniel Dennett, to embrace the *negation* of its conclusion. (Dennett's tendency to embrace the negation of the argument's conclusion incorrectly assumes, according to Lenzen, that a computational account of the role of emotion in cognition is possible, whereas Searle's original defence of the argument incorrectly assumes that a computational account of the role of perception in cognition is not possible.) Lenzen's interesting

arguments against both Searle and Dennett are clearly presented and make his paper a useful addition to a complex debate, though neither the complexity nor the current state of that debate are reflected in its pages. (Interested readers might wish to consult the recent collection, *Views into the Chinese Room* (Preston and Bishop 2002), which records the latest reflections on the topic both by Searle himself and by many of his most important critics.) Nida-Rümelin's paper ("Causal reduction, ontological reduction, and first-person ontology") is a critical examination of some central features of the account of consciousness that Searle has presented in *The Rediscovery of the Mind* (Searle 1992), and which he has subsequently used as a basis for his attacks on alternative approaches to the subject in *The Mystery of Consciousness* (Searle 1997a). The main critical target of the paper is Searle's claim that, since the first-person ontology of consciousness is a trivial by-product of definitional practices, there is no real tension between his thesis that consciousness is causally reducible to neurobiological processes and his thesis that consciousness is ontologically irreducible to neurobiological processes. After carefully distinguishing three separate claims that Searle conflates in his statement of the causal reducibility thesis, and having explained the reasons why causal reducibility claims are generally thought to be seriously undermined by arguments for the explanatory opacity of consciousness due to Chalmers (1996) and to Levine (2001), the paper argues that the ontological irreducibility thesis is both non-trivially true and actually inconsistent with all versions of the causal reducibility thesis, and that, therefore, the causal reducibility thesis must, contrary to what Searle intended, be abandoned. Among the high points of the discussion leading to that conclusion are, first, a convincing demonstration that Searle's denial that his position is vulnerable to the explanatory opacity arguments depends on an illegitimate confusion between questions of conceptual necessity and questions of nomological necessity, and, secondly, a painstaking reconstruction and refutation of Searle's argument that the first-person ontology of consciousness is a by-product of definitional practices. The paper is, in my view, one of the most penetrating among the available critical discussions of Searle's views on consciousness, and the most significant contribution to the 'Mind' part of the volume.

The remaining two papers in the 'Mind' part of the volume belong to the uncomfortable genre of colloquium papers that are only tenuously related to a colloquium's theme and merely use that theme as an excuse for publicising their authors' research agendas. The paper by Thomas Bartelborth and Oliver Scholz ("Understanding utterances and other actions") is primarily a vehicle of its authors' intention to argue that the unbridgeable gap that many philosophers have

claimed to perceive between explanations in the natural sciences and explanations in the social sciences is non-existent, and that the idea that both kinds of explanation aim at the maximisation of coherence is the key to their unification. The connection with Searle is purportedly established in the paper's closing pages, where his analysis of metaphorical utterances is briefly summarised and heavily praised (without regard to any of the numerous criticisms that it has received) as an analysis that, allegedly, can be easily reinterpreted in coherence theoretic terms. The authors' limited familiarity with Searle's work is evident from the fact that, although their primary purpose is to *rebut* arguments claiming that natural science explanations and social science explanations cannot be unified, they obviously ignore, and therefore do not discuss, the long series of original arguments that Searle has produced *in favour* of the thesis that the two types of explanation cannot, for *logical* reasons, be unified (see especially Searle 1991c). Since these arguments directly contradict the authors' unification proposal, not discussing any of them in a volume specifically devoted to Searle's philosophy makes it difficult to figure out in exactly what sense this paper is supposed to be a relevant contribution to the volume's topic. As for the notion of coherence that, irrespective of the volume's topic, the authors are interested in promoting as a kind of all-purpose conceptual problem-solver, it is too vaguely presented in the paper to allow any conclusions about its possible value, and the most detailed models of it that are presently available are left unexplored. (It is remarkable, for example, that, although Paul Thagard is named as the authors' authority on coherence, no use is made of the distinctions introduced or of the analyses offered in his main work on the subject, *Coherence in Thought and Action* (Thagard 2000), whose content is actually in tension with the authors' remarks on metaphorical interpretation as a procedure aiming at explanatory coherence, since, according to Thagard, what motivates metaphorical interpretation is primarily a search for analogical rather than for explanatory coherence.) It seems to me, then, that the Bartelborth and Scholz paper is of no obvious interest to students of Searle, and of only marginal interest to students of either coherence or metaphor, though it might possibly be of some interest to students of some of the other topics that it briefly discusses (*ceteris paribus* laws, interpretational charity, rationalising explanation, etc.)

I doubt, however, that serious students of *any* topic could develop an interest for the paper by Thomas Roeper ("The hidden algebra of the mind from a linguistic perspective"), which concludes the 'Mind' part of the volume. Roeper is a linguist working within the Chomskyan tradition, who appears to have been disturbed by the perceived implications of an important new argument against

computational theories of mind, much more radical than the Chinese Room Argument, that Searle first presented in Searle (1990) and later used in *The Rediscovery of the Mind* (Searle 1992). The relevant part of the argument is to the effect that since any physical phenomenon whatsoever admits of a computational description, and since no computational description assigns, as such, an intrinsic, as opposed to an observer-relative, property to the phenomenon it characterises, no account of mental (including linguistic) phenomena will be scientifically adequate if it ignores the neurobiological and restricts itself to the computational level. If this argument is sound, its implications for the self-image of much of cognitive science (and, of course, of Chomskyan linguistics) are obviously negative. Roeper, however, has chosen, instead of seriously confronting the argument (as others have attempted to do), to try to exorcise it by emitting an obviously uncontrollable series of irrelevant remarks, among which the most comical appear to derive from his idea that Chomskyan linguists need not be worried by the fact that some of their claims are incompatible with Darwinian biology because, according to Roeper, it can be established by moral (!) arguments that Darwinian biology is scientifically inadequate (!). We are given to understand, for example, that Darwinian biology has a serious problem as a theory because (allegedly) "the parent who slaps a child while muttering, 'you little monkey', has been influenced by Darwinist theory" (p. 224), and, even more dramatically, that the scientific validity of Darwinian biology is doubtful because the "concept of fitness" developed in one of its "inevitable" theoretical extensions has been (allegedly) responsible for such things as "the Third Reich or ethnic cleansing in Kosovo" (p. 224). We are also informed that the only form of biology worth having is the one in which the pre-Darwinian concept of teleology is restored as the central explanatory concept, and that the best contemporary attempts at resurrecting such a pre-Darwinian biology are embodied in "Carl Sagan's claims that there should be life in other planets" (p. 225) as well as in "the notion of Universal Grammar introduced by Chomsky" (p. 225). And we are finally given a list of urgent research problems that a fully resurrected pre-Darwinian biology should, according to Roeper, strive to solve, among which one of the most prominent is, it appears, the problem of "find[ing] a way to state that the musical properties of Mozart are part of the definition of fingertips" (p. 227). Subscribing to the author's statement that "dignity is the core of our sense of the self" (p. 223), I shall not comment on his paper any further. It seems to me, however, that the volume's editors would owe their readers a note explaining why and how their evaluation procedures have been relaxed in order to accommodate the paper in an otherwise serious volume.

The five papers in the 'Social Reality' part of the volume are all concerned with aspects of Searle's book *The Construction of Social Reality* (Searle 1995),

which constitutes his main contribution to date to the relatively new field of the analytic study of social ontology. Since all the papers express sympathy for the central motivating assumption of that book (the assumption, namely, that, because social reality exists only in so far as it is believed to exist, and therefore is not ontologically objective, the main philosophical problem that it raises is how objective knowledge of it is possible), it should perhaps be noted that the sympathy is by no means universally shared, and that readers wishing to acquire a more realistic view of the reception of Searle's work in this area would do well to consult additional sources. (A good starting point would be the collective volume on *John Searle's Ideas about Social Reality* (Koepsell and Moss 2003), which reprints recent critical discussions of Searle's social ontology, among which some—in particular, Viskovatoff (2003) and Fitzpatrick (2003)—express much deeper worries about Searle's project than any of the worries manifested by the present volume's contributors; among earlier discussions, Ruben (1997) and Wettersten (1998) should also be consulted.) In spite of their fundamental sympathy for Searle's project, however, the five papers in the 'Social Reality' part express several reservations, many of them justified, either about the way in which Searle attempts to embed his project into wider ontological disputes (in particular, the realist/anti-realist dispute), or about the conceptual resources that he utilises in order to implement it (in particular, about the three fundamental concepts—collective intentionality, status function assignment, and constitutive rule—that he uses in constructing definitions of social notions, and about the relation between these concepts and his concept of the Background).

David Sosa's very short paper ("True reality and real truth") concentrates on Searle's claim that physical, as opposed to social, reality is not belief-dependent, and notes that the way Searle uses that claim as a point of entry into the realism/anti-realism dispute suffers both from crucial ambiguities and from certain mischaracterisations of both the realist and the anti-realist positions in that dispute. Stanley Barry Barnes's paper ("Searle on social reality") offers a typically constructivist discussion of that same claim of Searle's, arguing that, although social reality is indeed, as Searle proposes, entirely belief-dependent, it is not the case that physical reality is, as Searle assumes, belief-*in*dependent. (Barnes seems either unwilling to admit or unable to realise that his arguments, if sound, lead to a form of generalised anti-realism, and that, since Searle certainly rejects any form of generalised anti-realism, his position is much more at odds with Searle's than he professes). Josef Moural's paper ("Searle's theory of institutional facts") gives a useful outline of the Searlian account of the subset of social facts that Searle labels 'institutional facts', argues convincingly that, on some important points of detail (for example, the use of negation in the representation of

events of institutional fact termination), Searle's account is inadequate, and assembles considerable textual evidence suggesting that there is a fundamental tension between Searle's explicitly *representationalist* analysis of the conditions of existence of institutional facts in terms of the collective assignment of status functions through constitutive rules and his explicitly *anti-representationalist* account of the conditions of existence of institutional facts in terms of the notion of Background. (Moural does not comment either on the sources or on the consequences of that tension; interested readers might wish to note that much the same tension has been independently detected, and richly commented upon from various perspectives, in four recent papers on Searle—Turner (1999); Dreyfus (2001); Viskovatoff (2002); Thalos (2003)—that virtually pick up the discussion at the point where Moural leaves it). Raimo Tuomela's paper ("Searle, collective intentionality, and social institutions") also comments on Searle's analysis of institutional facts, mainly from the point of view of Tuomela's own theory of social institutions, whose highly compressed exposition appears to be the main goal of the paper. (A more leisurely statement of that theory is available in Tuomela's book, *Philosophy of Social Practices* (Tuomela 2002), which is not cited in the paper, but is the source of most of its material.) Tuomela's two main criticisms of Searle—namely, that his account fails to recognise social institutions that are merely based on mutual expectations rather than on mutual agreements, and that it fails to recognise social institutions that merely generate conceptual rather than deontic powers—are not developed in great detail, but they do seem to identify genuine issues that Searle's account of social reality would, in the long run, have to face (even though Searle might conceivably attempt to postpone discussion of these issues, by claiming that the social phenomena that Tuomela considers are not properly called institutional, and should be given a different sort of treatment). Finally, Georg Meggle's paper ("On Searle's collective intentionality") compares and contrasts the way in which Searle defines the notion of collective intentionality in his account of social facts with the more standard ways in which that notion is defined in alternative accounts of social facts within analytic studies of social ontology. (Very roughly, a group's collective intentionality reduces, according to the standard accounts, to each group member's forming a certain 'I'-intention that is the object of mutual knowledge or belief, whereas, according to Searle, a group's collective intentionality reduces to each group member's forming a certain primitive 'we'-intention distinct from any 'I'-intentions, and not required to be the object of mutual knowledge or belief; for recent discussions of both positions, see the relevant contributions to the volume, *Social Facts and Collective Intentionality* (Meggle 2002)). Some of Meggle's critical remarks on Searle are tendentious. (For example, his claim that Searle's views can be demonstrated

to have the absurd implication that "any state of affairs is a social fact" (p. 265) depends on, among other things, the assumption that, according to Searle, when P is a social fact, the collective belief sustaining it is not necessarily the collective belief that P, but may be a collective belief whose content is wholly independent of P; however, Searle clearly rejects that assumption through his often repeated claim than, in order for a social fact to exist, a collective belief that *this very fact* exists is always necessary, and so Meggle's purported demonstration turns out to be directed against a view that neither Searle nor anyone else appears to have endorsed.) Nevertheless, some other critical remarks on Searle that Meggle makes are convincing. (For example, he correctly points out that if, as Searle claims, analyses of social facts in terms of mutually held 'I'-intentions have the *defect* of being infinitely regressive, then his own analysis of social facts in terms of *shared* 'we'-intentions has exactly the same defect, since it is also an analysis that, in order to be maintained, has to become infinitely regressive; and he, again correctly, points out that Searle's claim that social cooperation can only be analysed within a 'we'-based, and not within an 'I'-based, framework is unsuccessful, since it mistakenly assumes that 'I'-based approaches can only appeal to unconditionally and not to conditionally held preferences among group members.)

In general, the contributions to the 'Social Reality' part of the volume usefully identify specific problems that Searle's conception of social ontology should be acknowledged to have to resolve, assuming that its basic orientation is accepted as unproblematic. It is a pity, however, that none of the papers in the volume attempts to systematically examine the relations between Searle's late interest in social ontology and his long standing interest in speech acts, especially since a philosophical approach to speech acts much earlier than Searle's has persuasively been claimed—see Smith (1990) and the relevant papers in Mulligan (1987)—to have afforded insights unavailable to Searle's approach precisely because, unlike his approach, it was explicitly guided by considerations of social ontology from the very beginning.

Turning to the 'Speech Acts' part of the volume, I shall be brief on the papers by Daniel Vanderveken and by Frank Kannetzky, since neither of them leaves much room for comment. Vanderveken's paper ("Searle on Meaning and Action") is mainly concerned with summarising either Searle's well-known work in selected areas or Vanderveken's less well-known, but already published, work in the same areas, and its reliability as a summary is questionable. Even disregarding such solecisms as Vanderveken's reference to Austin's trichotomy of locutionary, illocutionary and perlocutionary acts as "Austin's trilogy [sic] of locutionary, illocutionary and perlocutionary acts" (p. 142), the coherence of the exposition is often quite doubtful. For example, echoing Searle's thesis that

"there are not, as Wittgenstein (…) and many others have claimed, an infinite or indefinite number of language games or uses of language" (Searle 1979: 29), Vanderveken asserts that "Searle challenged the anti-theoretical position of the second [=later, SLT] Wittgenstein according to whom there are uncountably many different kinds of language use" (p. 142); a few pages later, however, Vanderveken affirms that "Wittgenstein *and Searle* are right to say that it is impossible to construct a theory of all possible kinds of language games" (p. 158; italics added), citing as the basis of the anti-theoretical view that, in his view, Wittgenstein *and Searle* share Wittgenstein's contention that "there are 'countless kinds' of language games" (p. 158); since readers familiar with the relevant texts could easily determine that Searle does *not* have the contradictory attitudes toward Wittgenstein (or toward theory-construction) that these assertions represent him as having, it seems that the only thing that the assertions succeed in showing is that Vanderveken has, apparently, not yet decided whether he should side with Searle *or* with Wittgenstein on the question that clearly divides *them*; I doubt, however, that learning about Vanderveken's states of indecisiveness would be the primary preoccupation of prospective readers of the G&M volume.

As for Kannetzky's paper ("Expressibility, explicability, and taxonomy"), its professed purpose is to articulate a critique of Searle's so-called 'principle of expressibility' (Searle 1969: 19–21). It articulates no such critique, however, since it begins by radically misinterpreting the principle as a kind of recipe designed to give us practical help in dealing with situations where utterances "can be misunderstood because of hidden parameters" (p. 78), and then irrelevantly claims that we do not, in our everyday dealings as speakers and hearers, need the help that the principle allegedly purports to give us (because, in cases of communicative uncertainty, "we can", as hearers, "simply ask [the speakers] how an utterance is to be understood" (p. 79), and, as speakers, help our hearers understand by, for example, employing "analogies or metaphors", or by "using antonyms and negations [in the] hop[e] that the hearer knows them" (p. 78)). (I presume it is obvious that all this, besides being rather naïve, has nothing to do with Searle's principle of expressibility—which does not mean, of course, that, once properly elucidated, the principle cannot be thoroughly disputed; for an important recent elucidation and critique, see Récanati 2003; for interesting early discussions, ignored by Kannetzky, see Binkley 1979 and Gazdar 1981.)

Of the remaining six papers in the 'Speech Acts' part of the volume, three are devoted to Searle's declarational analysis of explicit performatives, and three discuss more general issues of the theory of speech acts that he has developed in *Speech Acts* (Searle 1969), *Expression and Meaning* (Searle 1979), and subsequent works. I will discuss the two groups of papers separately.

Searle developed his declarational analysis of explicit performatives in "How performatives work" (Searle 1989), perhaps the most frequently cited among the articles reprinted in his recent book of collected essays, *Consciousness and Language* (Searle 2002). (The first declarational analysis of explicit performatives, never cited by Searle but more thoroughly developed than his own, was presented eight years before Searle's article in a book by François Récanati (1981); the analysis was further developed in the expanded English version of that book (Récanati 1987), which Searle similarly never cites.) In his article, Searle assumes, following standard practice, that the distinguishing feature of a present tense explicit performative utterance of the form "I (hereby) $V(p)$" is that, in the right context, its utterance is sufficient for bringing into existence the illocutionary act that its speaker names under $V(p)$. He opposes, however, both those who (like himself at an earlier time) would take that feature to be a primitive one requiring no explanation, and those who (like himself at a different earlier time) would be tempted to explain it by claiming that the performance, by the speaker of an explicit performative, of the act named under $V(p)$ is a consequence of that speaker's additionally performing the unnamed act of *stating* that he is performing the act named under $V(p)$. This unnamed act of stating is indeed performed, Searle now says, but it is *itself* a consequence of, among other things, the performance of the act named under $V(p)$, and so cannot be invoked in explaining why the act named under $V(p)$ is performed; rather, the act named under $V(p)$ is performed, Searle now believes, as a result of the speaker's performing another unnamed act, the act of *declaring* that he is performing the act named under $V(p)$; and the task of an analysis of performativity is to show exactly how a speaker's declaring, in the right context, that he is performing an illocutionary act has as a *logical* consequence his performance, in that context, of that illocutionary act. The central elements of Searle's analysis are, first, that declaring that one performs a given act by means of a given utterance amounts to *manifesting the intention* to perform that act by that very utterance, and, second, that illocutionary acts have the special property that *recognising* a person's manifest intention to accomplish them *suffices* for that person's accomplishing them; assuming, then, that the literal meaning of "I (hereby) $V(p)$" is such that its speaker *declares* that he is performing a certain act named under $V(p)$, it follows—Searle contends—that, when the act that happens to be named under $V(p)$ is an *illocutionary* act, the speaker will automatically succeed in performing the act that he names.

The three papers on Searle's declarational analysis of explicit performatives (by Al Martinich, Robert Harnish, and Günther Grewendorf, respectively) are unanimous in rejecting the analysis, but only two among them offer any reasons justifying the rejection. The one that does not is the paper by Martinich ("On the

proper treatment of performatives"), whose allegedly critical remarks on Searle belong to the one or the other of two categories: those in which he simply repeats the analysis and then urges us to abandon it just on the grounds that its claims strike him as "completely counter-intuitive" (p. 98) or "unintuitive" (p. 99), and those in which he actually tries to refute the analysis by arguments that only succeed in revealing his incomplete understanding of it. Two examples of the latter category will have to suffice. In voicing his preliminary reservations about Searle's approach, Martinich claims that it does not do justice to the (alleged) fact that "*Leave the room* is intuitively as performative as *I order you to leave the room*" (p. 94). However, the question whether or not these two utterances are both 'performative' is not a matter of intuition but of definition. On a now rarely used sense of the term 'performative', they are both performative, for the trivial reason that, in *that* sense, *every* serious utterance is performative (because, in *that* sense, an utterance is performative just when its speaker performs some illocutionary act, whether or not he *self-ascribes* that act). But on the much narrower, and far from trivial, sense of 'performative' that has come to predominate (the sense in which an utterance is performative when its speaker not only performs an illocutionary act but also *self-ascribes* that act, and which is the *only* sense in which the terms 'performative' and 'explicit performative' are synonymous), the first of the above utterances is certainly not performative whereas the second certainly is. And since Searle, at the very beginning of "How performatives work", does distinguish between the two senses and does state that it is exclusively with performatives *in the narrow sense* (in other words, with explicit performatives) that his analysis is concerned, it is hard to see how anyone's habit of using the term *in the wide sense*, which would in any case trivialise the notion of performativity, could count as an argument against Searle's approach. In what appears to be his central objection to Searle, Martinich claims that the declarational analysis of explicit performatives is unsuccessful because it destroys the coherence of the Searlian taxonomy of illocutionary acts, which Martinich unreservedly accepts and affectionately calls "the good old theory" (p. 99). His reasoning here is that, since illocutionary acts of all kinds can be performed by means of explicit performative utterances, and since, on the declarational analysis, explicit performatives perform unnamed acts of declaration, all kinds of illocutionary acts would, on the declarational analysis, turn out to be subcategories of the category of declarations. Martinich calls this alleged result of the declarational analysis an "absurdity" (p. 99). However, no such result can be derived from the declarational analysis, except by someone who commits several elementary mistakes, among which the most obvious is that of interpreting the fact that an act A can be a *means* for the realisation of an act B as evidence that B-type acts are *species* of A-type

acts: Just as the fact that one can express a categorical denial by uttering a rhetorical question does not make denials a species of questions, so the possibility of realising illocutionary acts of all kinds by means of explicit performatives that constitute declarations does not entail that illocutionary acts of all kinds thereby become species of declarations. Martinich's central argument against Searle commits that basic mistake, which is really not different in kind from the mistake of supposing that, since one can cause someone to wake up or to fall asleep by singing songs to them, the act-types 'causing someone to wake up' and 'causing someone to fall asleep' are *species* of the act-type 'singing songs to someone'. But since it is only Martinich and not Searle who commits that mistake, neither Searle's declarational analysis of explicit performatives nor his taxonomy of illocutionary acts are in any way threatened by anything Martinich says about them.

For serious critical discussion of Searle's declarational analysis, readers of the volume should turn to the papers by Robert Harnish and Günther Grewendorf, who in effect defend, respectively, the two kinds of position concerning explicit performatives that Searle implicitly or explicitly discards before arriving at the declarational analysis. Harnish's paper ("Are performative utterances declarations?") claims that the hypothesis, articulated in Bach and Harnish (1979), that speakers of explicit performatives perform the acts they name by way of performing unnamed acts of *stating* provides a better overall account of the phenomena than Searle's hypothesis that they perform the acts they name by way of performing unnamed acts of *declaring*. Harnish supports that claim by arguing (a) that the specific objections raised by Searle against the hypothesis of Bach and Harnish (1979) are not valid, and (b) that Searle's own hypothesis, to the extent that its content is clear, rests on dubiously accurate or dubiously consistent assumptions. Some of Harnish's objections to Searle's account of performatives require acceptance of the Bach and Harnish (1979) account, and might, for that reason, fail to convince uncommitted observers. (Such observers might note, for example, that Searle's account, for all its problems, purports to cover *all* explicit performatives, whereas the Bach and Harnish (1979) account begins by *excluding* from consideration all kinds of explicit performatives naming so-called 'essentially conventional' illocutionary acts (for the original distinction between 'essentially conventional' and 'not essentially conventional' illocutionary acts, see Strawson 1964); that exclusion, however, makes it highly doubtful whether the Bach and Harnish (1979) account, even if it was unobjectionable within its restricted domain, could be regarded as an *alternative* to any account which, like Searle's, purports to cover *all* explicit performatives, especially since the "conditions of adequacy" on accounts of performatives that Harnish outlines at the beginning of his paper nowhere stipulate that adequate accounts should *exclude*

performatives naming so-called 'essentially conventional' illocutionary acts.) However, many of Harnish's arguments against Searle do not presuppose any thesis of Bach and Harnish (1979), and the one among them that, to my mind, is the most decisive (namely, the argument to the effect that the declarational analysis cannot consistently maintain its claim that performativity is a consequence of literal sentence meaning and its claim that performative sentences are not ambiguous between a performative and a non-performative reading) shows compellingly that, if the declarational analysis is to survive, it will definitely have to repudiate the non-ambiguity thesis with which it was explicitly associated by Searle. Besides, Harnish's close attention to Searle's relevant texts reveals that, on many questions that would be crucial to its evaluation (for example, the question as to how Searle's distinction between linguistic and extra-linguistic declarations should be validated), the declarational analysis supplies either no clear answers or no answers whatsoever and that, consequently, its eventual survival would depend not only on the repudiation of theses with which it was explicitly associated but also on the elaboration and defence of distinctions that it assumes but does not justify. For anyone interested in Searle's declarational analysis, then, Harnish's paper would be essential reading, independently of what one's opinion happens to be on the viability of the analysis that Bach and Harnish (1979) propose for the *subset* of performatives naming so-called 'not essentially conventional' illocutionary acts.

Grewendorf's paper ("How performatives don't work") seems attracted to the idea that performativity is a primitive phenomenon requiring no explanation, although that idea is not explicitly stated but rather suggests itself through the author's proposed elimination of both of the theses on which the current explanatory competitors respectively depend (that is, the thesis that speakers of explicit performatives perform unnamed acts of stating, and the thesis that they perform unnamed acts of declaring). The first of the two main arguments that Grewendorf advances against both theses (and which draws on an observation made by Schiffer 1972 long ago, with different aims) is, in effect, the following: If explicit performatives realised, besides the acts they name, unnamed acts of stating or of declaring, then (for reasons suggested by Schiffer 1972) their *full* illocutionary force could not be made explicit; but explicit performatives do make their full illocutionary force explicit; therefore, explicit performatives do not realise unnamed acts of stating or of declaring. This argument is unsuccessful, since it equivocates on the interpretation of 'full illocutionary force' and clearly begs the question against the views that it purports to rebut: when Grewendorf affirms, in the second premise, that each explicit performative makes its full force explicit,

he simply assumes without argument that the only force that an explicit performative carries is the one *that it names*; but that is precisely what the theorists he wants to oppose dispute, and it is, in any case, only by taking the full force of explicit performatives to include *unnamed* components that one would have any tendency to accept, *via* Schiffer (1972), what the first premise conditionally affirms, namely, that the full force of explicit performatives, precisely *because* it always includes unmanned components, cannot be made fully explicit. The second of the two main arguments that Grewendorf advances against the currently competing explanations of performativity is more interesting. Grewendorf notes that all these explanations require interpreting the main verbs of explicit performatives as indicators of propositional content rather than of illocutionary force, and they thus claim to be able to make clear theoretical sense of the idea that what speakers of explicit performatives *say* is that they perform certain acts that they do in fact perform; according to Grewendorf, however, the only sense in which a speaker of an explicit performative *says* that he performs a certain act that he performs is Austin's *phatic* sense of 'say', which is irrelevant to the determination of propositional content, and not Austin's *rhetic* sense of 'say', which would be relevant to the determination of that content; it would seem to follow, then, that all currently competing explanations of performativity rest on incorrect, or at least wholly unmotivated, assumptions as to what the propositional content of explicit performatives is. This interesting argument would require, in order to be convincing, much more development than it actually receives in the paper. For one thing, the author should at least mention a well-known paper by Searle (1968), where Austin's notion of a locutionary act, relative to which the phatic/rhetic distinction is defined, is rejected. For another thing, and most importantly for his purposes, he should both mention and use several studies that have tried either to defend Austin's notion of a locutionary act (and so, the phatic/rhetic distinction) against Searle's objections—see especially Thau (1972), Forguson (1973), Frye (1976)—or to explore various (and not always congruous) directions in which Austin's unsystematic statements about locutionary acts (and their phatic and rhetic components) could be developed in order to become reliable analytical tools—see, for example, Griffiths (1969); Furberg (1971); Strawson (1973); Holdcroft (1978); Récanati (1980). Even in the absence of such developments, however, Grewendorf's argument points to a real issue whose examination is largely neglected in current discussions of explicit performatives, and might turn out to be important in evaluating purported explanations of them.

Though a thorough examination of Searle's declarational analysis would obviously be out of place here, let me conclude this part of the discussion by noting two important, and—as far as I know—previously undiagnosed, dilemmas that

Searle's conception of performatives yields when combined with certain *other* elements of his philosophy of language. The first dilemma is a consequence of the assumption of the declarational analysis that the main verbs of explicit performatives are never indicators of illocutionary force and always indicators of propositional content. That assumption is in direct conflict with Searle's proposal to distinguish between two kinds of semantically relevant operators on sentences, the illocutionary and the propositional ones, since that distinction (which is supposed by him to be logically fundamental) is impossible to maintain *except* on the assumption that the main verbs of explicit performative sentences are always indicators of illocutionary force and *never* indicators of propositional content. (For example, saying, as Searle would recommend, that the negation in *I promise that I won't come* is propositional whereas the negation in *I don't promise that I will come* is illocutionary presupposes that, in *I promise that I will come*, the performative verb is an indicator of illocutionary force and *not* an indicator of propositional content; but that presupposition is inconsistent with the declarational analysis of *I promise that I will come*, where the performative verb has to be taken to specify the content and *not* the force of the unnamed act of declaration that is allegedly being performed. Similarly, saying, as Searle would recommend, that the conditional in *I predict that, if you want my opinion, he will win* is propositional, whereas the conditional in *If you want my opinion, I predict that he will win* is illocutionary presupposes that, in *I predict that he will win*, the performative verb is an indicator of illocutionary force and *not* an indicator of propositional content; but this is inconsistent with the declarational analysis of *I predict that he will win*, where the performative verb has to be taken to specify the content and *not* the force of the unnamed act of declaration that is allegedly being performed.) The first dilemma, then, is that Searle must abandon *either* the particular application of the force-content distinction demanded by his declarational analysis of explicit performatives (and so the declarational analysis itself) *or* his distinction between illocutionary and propositional interpretations of sentence-forming operators (and so, whatever he thought was the fundamental logical insight afforded by that distinction). The second dilemma derives from the declarational analysis' claim that it succeeds in representing the performativity of explicit performatives as a logical consequence of their literal meaning, by assuming that, in virtue of that literal meaning, their speakers manifest intentions (to perform certain acts) that are guaranteed to be fulfilled as soon as they are *recognised* to be present. That claim is in direct conflict with the particular kind of intentionalist theory of meaning that Searle has developed (see Searle 1986a and Chapter 6 of Searle 1983) in opposition to Grice's intentionalist theory

(see Grice 1957 and the relevant papers reprinted in Grice 1989), since what crucially distinguishes the two theories from each other is precisely that, on the Gricean theory, meaning intentions are essentially communicative in that they are never fulfilled unless they are recognised to be present, whereas, on the Searlian theory, meaning intentions are not essentially communicative since their fulfillment is independent of whether or not they are recognised to be present. Clearly, however, if, as Searle contends against Grice, the fulfilment of meaning intentions is always recognition-*independent*, it cannot be the case, as the declarational analysis affirms, that the performativity of explicit performatives derives *logically* from the recognition-*dependent* fulfilment of the meaning intentions they convey. The second dilemma is, then, that Searle must *either* abandon the claim that the performativity of explicit performatives is a logical consequence of their meaning (which was the most distinctive claim of his declarational analysis) *or* the claim that the fulfilment of meaning intentions is always recognition-independent (which was the most distinctive claim of the intentionalist theory of meaning that he has developed in opposition to Grice's intentionalist theory). Exactly how, if at all, these dilemmas can be resolved is a complex question requiring separate treatment. But the very fact that they arise shows clearly, I submit, that, far from being a natural development of Searle's conception of linguistic meaning, the declarational analysis is actually a proposal that forces many of the tensions inherent in that conception to come to the surface.

The remaining three papers in the 'Speech Acts' part of the volume (by Andreas Kemmerling, Mark Siebel, and Christian Plunze, respectively) move beyond the topic of performatives, and two among them subject certain aspects of Searle's philosophy of language to some of the most searching critical scrutiny that they have received in the recent literature. The paper by Andreas Kemmerling ("Expressing an intentional state") is not critical of any thesis of Searle's (if one excludes an incidental remark dismissing "the whole thing about so called constitutive rules" (p. 83)), but concentrates on an unanalysed notion that Searle often uses, and attempts to offer an analysis of it. The notion in question is that of a person's *expressing* a mental state without necessarily *being* in that state, and is frequently used by Searle in his accounts of various kinds of speech act, in particular expressive speech acts. Presumably because Kemmerling finds the notion interesting in its own right, he is not concerned with the fact that the Searlian accounts of speech acts employing it are multiply flawed (for some of the main reasons, see Tsohatzidis 1993b). It is best, therefore, to consider the analysis independently of its past uses or abuses within speech act theory, which is what Kemmerling himself does when presenting it. Very briefly, the proposed analysis is to the effect that a person X *expresses*, by behaving in way W, a mental state Y

even without *being* in that state Y, if, and only if, (1) X is a *normal* member of a population P and behaves in way W under *normal* circumstances, and (2) it is not synthetically but *analytically* true that, if a normal member of P behaves in way W under normal circumstances, then his behaving in way W constitutes *evidence*, within P, of his being in the state Y. I suspect that the analysis would be of no interest to those who would refuse to swallow the unanalysed concept of normalcy on which it essentially relies (and which it applies indiscriminately to either persons or circumstances), or to those who, for familiar reasons, would deny that a clear distinction between analytic and synthetic truths is possible. But even those who would not refuse to make these admittedly large concessions might find the analysis ultimately unsatisfactory. Suppose, for example, that there is a population R whose members live their entire lives under a dictatorial form of government, and that, simply in order to avoid persecution, *all* members of the population, in *all* sorts of circumstances where the question arises, make it a point to *express* the belief that their government is democratic, even though not a single one among them really *has* that belief; on Kemmerling's analysis, it would follow that it is a *conceptual truth* that the behaviour of the population provides *evidence* to every member of R that every other member of R really believes that the government is democratic; but since it can hardly be a conceptual truth that no distinction can possibly be made between epistemic justification and propaganda, it seems that Kemmerling's analysis is mistaken.

The paper by Mark Siebel ("What is an illocutionary point?") begins by noting that, although the notion of illocutionary point is fundamental to Searle's theory of speech acts, Searle has never supplied a definition of that notion, and proposes to examine how such a definition could be constructed on the basis of the definitions of the five illocutionary points underlying Searle's taxonomy of speech acts (Searle 1975a), and later used as foundations of his system of illocutionary logic (Searle and Vanderveken 1985). Siebel then shows that no less than four non-equivalent definitions of the notion of illocutionary point can be constructed on that basis, and, furthermore, that each one of these four definitions, when taken in conjunction both with the Searlian taxonomy's stated objectives and with the Searlian analyses of individual illocutionary acts, turns out to be problematic, either because it has the effect of making the proposed analyses of individual speech acts internally inconsistent or because it has the effect of violating the mutual exclusivity requirement on taxonomic classes that the taxonomy was aiming to respect. The author summarises his investigations by expressing the (surely understated) opinion that "it is hard to grasp the point of illocutionary point" (p. 138), and the arguments on which his conclusions are based are carefully stated and invariably compelling. Furthermore, the various

critical remarks on Searle's taxonomy that he makes on his way to these conclusions are all apposite, and, when combined with the results of some powerful independent critical studies of that taxonomy that he does not mention (in particular, Sadock 1994), suggest that the confidence with which Searle assumes the taxonomy's correctness in both informal and formal expositions of his theory of speech acts is unwarranted.

Finally, the paper by Christian Plunze ("Why do we mean something rather than nothing?") is a critical analysis, based on an examination of the special but central case of assertive speech acts, of Searle's attempt, already referred to above, to replace Grice's intentionalist theory of meaning with a new intentionalist theory, where the only meaning-constitutive intentions are not Gricean communicative intentions but rather what Searle calls 'representing intentions' (that is, intentions whose fulfilment is recognition-independent, and whose goal is to impose on not intrinsically intentional physical objects—namely, utterances—conditions of satisfaction held under various illocutionary modes). The author's main conclusion is that Searle has failed to demonstrate that his intentionalist theory is superior to Grice's. That conclusion is based on two series of arguments: (a) those in which Searle's criticisms of Grice are claimed to be ineffective either because they are based on counterexamples that are not genuine or because they are based on counterexamples that, though genuine, are also counterexamples to Searle's own approach; and (b) those in which Searle's positive account of meaning in terms of 'representing intentions' is claimed to be inadequate either because it fails to preserve obvious distinctions or because it has absurd consequences. The arguments are persuasively deployed, and those, in particular, that are directed against Searle's notion of 'representing intention' belong to the best contemporary discussions of the idiosyncratic version of intentionalism that Searle has tried to develop in the theory of meaning. Indeed, when taken in conjunction with the results of two highly relevant critiques, which Plunze does not mention, of Searle's anti-Gricean arguments (Bennett 1991 and Siebel 2001), they make it highly plausible to assume that, *if* an intention-theoretic account of meaning was to be adopted at all, then the Gricean conception of meaning-intentions would be preferable to the Searlian one as a basis of such an account.

To summarise all the preceding remarks, my opinion on the G&M volume is, on the whole, positive. Although six out of the eighteen papers that constitute its core suffer from various kinds of serious weaknesses, the remaining twelve papers make interesting, and in some cases important, contributions to the study of their topics. And although neither Searle's unwillingness to contribute a written reply to his critics nor the editors' unwillingness to contribute an introduction to the critics' arguments make the volume an especially impressive sequel to the

Lepore and Van Gulick (1991) collection of critical essays on Searle's philosophy, it will be a valuable resource for those interested in all three of the main areas of Searle's philosophical work, especially if used in conjunction both with that earlier collection and with a more recent collection of critical essays on Searle (Smith 2003), which is comparably broad in scope.

3 Remarks on V&K

The editors' introduction to the V&K volume provides brief summaries of the volume's papers, but what it attempts to do beyond that—namely, to offer a "historic [sic] survey" (p. 1) of speech act theory, and to tell readers where "the future of speech act theory lies" (p. 18)—should perhaps have been left for another, more mature, occasion. The "historic" [sic] survey turns out to be a catalogue of problems in the theory of meaning that, allegedly, have been fully and definitively solved in a previous book by one of the editors (even though they "were completely ignored" (p. 10) by such figures as Aristotle, Frege, Russell, Carnap and Davidson, to mention but a few); and the direction in which "the future of speech act theory lies" is announced to be identical to the direction recommended by the same editor's "recent papers" (p. 18)—which, though *not* included in the volume, are described in the introduction much more thoroughly than any of the papers that the volume does include. In short, the introduction's primary purpose appears to be to alert the world to the presumed significance of the work on various topics that one of the volume's editors has done in the past; but since that work, whatever its value, has been done *elsewhere*, it is unclear why the volume's readers should be supposed to be unable to absorb the volume's contents without antecedently becoming convinced of that work's presumed significance. One fact that the introduction does succeed in making clear, no doubt unintentionally, is that readers should perhaps be prepared to encounter many instances of editorial carelessness throughout the volume. To give one among many examples: Although the introduction is presented as co-authored by the volume's *two* editors, it has a footnote reading, "I am grateful to [x] for [y]" (p. 285), which is uninterpretable except on the metaphysically extravagant assumption that fusions of individuals are possible.

Besides the editorial introduction, the volume contains twelve papers, which, with one exception, have not been previously published. The exception is John Searle's paper "How performatives work", published fifteen years ago (Searle 1989), whose reprinting in the volume is difficult to understand, not only because it is a well-known paper easily accessible elsewhere, but also because

none of the other papers in the volume enters into any kind of systematic dialogue with it. Of the remaining eleven papers, the one by Daniel Vanderveken is by far the longest in the book. The other ten papers are for the most part thematically disparate and, in the majority of cases, far more narrowly focused than their titles suggest. I will briefly comment on them, following (in the absence of any obviously applicable principle of categorisation) the order in which they appear, and will return to Vanderveken's contribution before the conclusion.

The paper by André Leclerc ("Verbal moods and sentence moods in the tradition of universal grammar") is a largely doxographic account of the not always consistent ways in which the question of the semantics of verbal mood was treated in the Port Royal logic and the Port Royal grammar, as well as a description of the different forms in which different elements of the Port Royal tradition have survived in the work on verbal mood by various grammarians-philosophers of the European Enlightenment. No systematic attempt is made either to elucidate the intellectual contexts within which the individual views described in the paper have been developed or to connect these views to contemporary issues in the analysis of mood. Indeed, some significant recent research that has connected certain Enlightenment conceptions of mood to specifically speech act theoretic issues is not even mentioned. For example, Leclerc expresses the opinion that James Gregory's account of mood "is the most advanced and interesting of all those developed during the Enlightenment" (p. 78). But since the originality of that account is largely due to the fact, noted by Gregory himself and acknowledged by Leclerc, that it incorporates Thomas Reid's distinction between 'solitary operations of the mind' and 'social operations of the mind', one might expect that Leclerc would grasp this opportunity in order to investigate in some detail the use of that distinction in Reid's own philosophy of language. No such investigation is attempted, however, and no reference is made to a revealing study (Schuhmann and Smith 1990) that both undertakes such an investigation and shows in some detail that, by insisting that what distinguishes a 'solitary' from a 'social' operation of the mind is that the former can whereas the later cannot exist without being communicated, Reid came close to arriving at what is recognisably a distinctive contemporary position in speech act theory. One might expect that the paper would at least be serviceable as a source of information on aspects of the history of the study of mood for readers who happen not to know French; it cannot, however, be recommended even for that limited purpose, first because none of its numerous, and sometimes lengthy, quotations from French is translated, and secondly because the author's command *of English* is often quite uncertain (as becomes evident from, for example, the use of "What does mean the view that [x]?" (p. 70) instead of 'What does the view that [x] mean?'; or of "a

respectable tentative [sic] to solve some of the problems" (p. 84) instead of 'a respectable attempt to solve some of the problems').

The paper by Candida Jaci de Sousa Melo ("Possible directions of fit between mind, language, and world") devotes most of its nine pages to repeating, on the one hand, Searle's well-known four-fold classification of possible 'directions of fit' between linguistic representations and extra-linguistic reality (Searle 1975a), and, on the other, Searle's claim that his attempted application of the 'direction of fit' apparatus to the analysis of the relation between mental representations and extra-mental reality has revealed an asymmetry between illocutionary acts and mental states, since, although there are certain illocutionary acts, namely declarations, that impose on their contents the so-called 'double' direction of fit, there are no mental states that impose on *their* contents the 'double' direction of fit (Searle 1983). The author does not dispute any of Searle's claims (and appears to be unaware of an important study by Humberstone (1992) that is highly critical of the ways in which Searle and others have sought to explicate the notion of 'direction of fit'), but announces that she has made a discovery that to some extent reduces the asymmetry that Searle had noticed—the alleged discovery being that, although there are indeed no mental *states* that impose on their contents the 'double' direction of fit, there are certain 'acts of thought' that impose on *their* contents the 'double' direction if fit. The reader who might at this point have become eager to learn what these 'acts of thought' are, and how Searle could have overlooked them, is informed that the 'acts of thought' in question are "illocutionary acts of declaration such as inaugurations, appointments, decrees, benedictions, confirmations, definitions [etc.]" (p. 111). But these, the reader might now feel inclined to protest, are the very acts that Searle has *already* described as imposing on their contents the 'double' direction of fit, *without* thereby feeling obliged to modify his asymmetry thesis, since, on his account, illocutionary acts are *not* 'acts of thought'. Where exactly, then, does the author's discovery lie? The 'discovery', it transpires, is that, according to the author, and contrary to Searle, illocutionary acts *are* 'acts of thought' ("illocutionary acts are conceptual acts of thought" (p. 111)), and that, therefore, since illocutionary acts of declaration have been described by Searle as imposing on their contents the 'double' direction of fit, they should also be described by him as 'acts of thought' imposing on their contents the 'double' direction of fit. The utter triviality of that 'discovery' has obviously failed to amaze the author; her readers, however, are guaranteed to be amazed both by the reasoning that has made the alleged discovery possible and by the implications she has managed to derive from it. Thus, the author's *only* apparent reason for claiming that illocutionary acts are 'acts of thought' is that "when we perform them we think" (p. 111)—which is about as cogent an argument

as the argument that, for example, the act of robbing a bank or the act of escaping from prison are 'acts of thought', since, presumably, 'when one performs them one thinks'. And one of the main implications that the author has derived from the 'discovery' that declarations are 'acts of thought', and that, therefore, it is immaterial whether or not their occurrence is verbally signalled, is that "nothing prevents us from enriching by declaration the mental 'language of our [sic] thought' (Fodor 1975)" (p. 114)—a claim that, among other things, manifests either complete ignorance or complete misunderstanding of the work cited, since, according to Fodor, if there is such a thing as a 'language of thought', it must be *innate*, and so it cannot be either 'enriched' or 'impoverished', no matter how many declarations, verbal or 'mental', anyone performs. The overall impression that this paper conveys is that of a mediocre essay by a beginning philosophy student (though some remarks, such as the one according to which smiling is a "propositional attitude" (p. 117) would appear to point to a region well below mediocrity, even by student standards). Readers should therefore not be surprised to learn, from the 'Notes on contributors' section at the end of the book, that the author is in fact a student working under the supervision of one of the volume's editors (p. 321); they might well be surprised, however, at the supervisor's decision to burden the volume with an apparently unmarked student essay.

Although the paper by Alain Trognon ("Speech acts and the logic of mutual understanding") is ostensibly written in English words, it contains many word combinations that are not English sentences, as well as many word combinations that are not interpretable English sentences (for example, "an assertive is verified in a conversation if the hearer considers it as true of its objective truth" (p. 133)), and could relatively safely be recommended only to readers who happen to know French and are likely to have the ability to identify the French expressions that Trognon was trying to translate by putting together English words (thus, a reader who just knows English will not be able to figure out what is meant by the word string "for this reason which it is not possible to prove that the truth of [x] is a consequence of [y]" (p. 133); however, if a reader happens to know how to interpret the French phrase 'pour cette raison qu'il n'est pas possible de prouver que la vérité de [x] est une conséquence de [y]', then he or she will probably be able to understand that what Trognon was trying to express by composing his word string is simply what would be expressed by the English phrase 'because it is not possible to prove that the truth of [x] is a consequence of [y]'). Turning to the paper's substance (to the extent that its substance is discernible in spite of its form), it appears that Trognon's main concern is to show that a certain 'law' of illocutionary logic can, on the assumption that it has been internalised by speakers and

hearers, be invoked in order to explain the interpretative phenomena usually discussed under the heading of 'indirect speech acts'. It turns out, however, that Trognon's understanding of illocutionary logic is in no better shape than his understanding of English, and that his proposed application of the 'law' to the few examples he considers makes no explanatory sense. To take two of these examples, consider the situation (call it S1) where a speaker A says *Do you know where today's newspaper is?* and his hearer B reacts by saying *I will get it*; or the situation (call it S2) where a speaker A says *There is a draft* and his hearer B reacts by closing a nearby window. The familiar problem that these examples pose is that of explaining how, by the time B's verbal or non-verbal reaction becomes available in S1 or S2, both A and B would be likely to presume that B has interpreted A's utterance as a request (to get the newspaper or to close the window, respectively), in spite of the fact that A's utterance in S1 or A's utterance in S2 do not linguistically encode the relevant requests. Trognon correctly perceives that (rhetoric to the contrary notwithstanding) conversation analysis is not really in a position to supply an explanation of these facts. However, his own explanation is both indefensible in its presumption to have attained its goal and deeply confused about the means it invokes for attaining that goal. The proposed explanation appears to be this: What makes A's utterance in S1 interpretable as a request for the newspaper, or A's utterance in S2 interpretable as a request for closing the window, is, first, that there is a law of illocutionary logic (presumably known to every speaking human) to the effect that requests, along with certain other types of speech acts, are such that they cannot be satisfied (that is, complied with) unless they are successful; and, second, that, since B's reactions in S1 and S2 *would* satisfy certain requests *if* those requests *were* made, the speaker and hearer of S1 and S2 are entitled to presume that these requests *have* been made. Thus, since B reacts in S1 in a way that *would* satisfy a request for getting the newspaper, *if* that request *were* to be made by A, A and B are both entitled to conclude that A *has* actually made that request; and since B reacts in S2 in a way that *would* satisfy a request for closing the window, *if* that request *were* to be been made by A, A and B are both entitled to conclude that A *has* actually made that request. Now, even superficial acquaintance with the system of illocutionary logic that Trognon uses (Searle and Vanderveken 1985) would be sufficient for realising that he simply does not understand, among many other things (too many to be detailed here), what "satisfaction" is supposed to mean in that system, since *his* notion of satisfaction, unlike the notion of satisfaction used in that system, absurdly entails that *both* a hearer who closes a window because he was requested to do so *and* a hearer who merely *says* that he will get the newspaper that he was requested to get 'satisfy' their respective requests. According to both illocutionary logic and

common sense, however, only a hearer who actually *does* what he is requested to do, and not a hearer who merely *says* he will do what he is requested to do, counts as having satisfied—that is, complied with—a request that has been addressed to him. The major problem with Trognon's proposed explanation, however, is independent of his confused understanding of illocutionary logic (which is actually irrelevant to his concerns), and would persist even if we agreed, for the sake of argument, on a disjunctive definition of 'satisfaction', according to which a hearer 'satisfies' a request just in case he *either* does what he is requested to do *or* merely says that he will do what he is requested to do. Assuming that interpretation, Trognon's explanation boils down to this: The reason why A's utterances in S1 and S2 are interpreted as conveying certain requests that they do not linguistically encode—in other words, the reason why they are interpreted as conveying certain indirect requests—is that B's verbal or non-verbal reactions in S1 and S2 are such that they *would* 'satisfy' (in the disjunctive sense of 'satisfaction') these requests, *if* these requests *were* made. The explanation is totally empty, however, as can be seen from the fact that it fails to specify either necessary or sufficient conditions for the occurrence of the phenomenon to be explained. To see that no necessary condition is specified, consider a situation S3 where A says *Do you know where today's newspaper is?* and B reacts by saying *Don't even think I shall get today's newspaper for you*, or a situation S4 where A says *There is a draft* and B reacts not by closing, but by opening even more widely the window that is obviously implicated in the occurrence of the draft. It is clear that in these situations A's utterances would *still* be interpretable as indirect requests (for getting the newspaper or for closing the window, respectively), even though B's reactions could not possibly be interpreted as *satisfying* these requests, even in the disjunctive sense of 'satisfaction'. Trognon's explanation, therefore, does not even succeed in specifying a necessary condition on the interpretability of an utterance as an indirect request. And it is clear, of course, that it does not specify a sufficient condition either. Consider, for example, a situation S5 where a speaker A utters the sentence *It's a nice day today*, and her hearer B reacts by slapping the speaker across her face. Since it is obvious that the sentence *It's a nice day today* does not linguistically encode a request for a slap, and since it is also obvious that slapping a person across her face *would* 'satisfy' that person's request for a slap, *if* that request had been made, it ought to be the case, if Trognon's explanation specified a sufficient condition, that both A and B would readily interpret the utterance of *It's a nice day today* in S5 as an indirect request for a slap. Since this is evidently not the case, it appears that the reasons that really account for the fact that certain utterances can, and certain other utterances cannot, be interpreted as indirect requests are totally invisible to Trognon's 'explanation'.

The paper by Steven Davis ("Utterance acts and speech acts") is not primarily a contribution to speech act theory, but rather an attack on Donald Davidson's views on the individuation of events and actions (see the papers collected in Davidson 1980), to the extent that these views imply that, just as, when the turning of a key amounts to the locking a door, a single event (describable in two different ways) takes place, so, when the performance of an utterance act amounts to the performance of an illocutionary act, a single event (describable in two different ways) takes place. (Notice that Davidson's view is only a view about token-identities, not a view about type-identities.) Davis contends that, by mimicking an argument that Tyler Burge (1979) has long ago produced against the thesis that token mental states are identical to token brain states, he can construct an effective argument against Davidson's implied view about the token-identity of utterance acts with illocutionary acts. To roughly summarise, Burge's (1979) argument invites us to consider a given individual X with respect to two possible words, W1 and W2, of which the first is the actual world and the second is identical to the actual world *except* for the fact that what is H_2O in W1 has been replaced in W2 with a substance totally different chemically from, though phenomenologically identical to, the substance that, in W1, is H_2O; Burge then asserts that a sentence like *X believes that water is good to drink* might be true of X relative to W1 but could not be true of X relative to W2; and he concludes that token brain states are not identical to token mental states, since X's brain states would, by hypothesis, be the same in W1 and W2, whereas X's reported belief would be true in W1 and false in W2. Davis's variation on Burge's argument (which was itself a variation on a popular Putnamian argument), is that, in the setting described by Burge, an utterance act report like *X said "Water is good to drink"* would, if true of X, be true of X both relative to W1 and relative to W2, whereas an illocutionary act report like *X said that water is good to drink* might be true of X relative to W1 but could not be true of X relative to W2; from which Davis concludes that utterance acts cannot, contrary to Davidson's implied view, be token-identical with illocutionary acts. The argument is not compelling, however. For one thing, even those who do accept appeals to possible worlds as legitimate in philosophical argumentation would have good reason not to accept Davis's central claim. For example, assuming that the adjectives "real" and "earthly" are not meaningless (which they certainly aren't), and supposing that an inhabitant X of world W2—which, by hypothesis, contains no H_2O—has produced a token of "Water is good to drink", the conjunctive sentence *X said that water is good to drink, but it is not to real/earthly water that he was referring* would be *true* rather than false relative to W2; but this is clearly inconsistent with Davis's claim that the sentence *X said that water is good to drink* could *never* be true relative to W2, and suffices for

blocking his anti-Davidsonian conclusion. For another thing, Davidson has repeatedly indicated his reasons for regarding appeals to possible worlds as illegitimate in philosophical argumentation, and it is therefore pointless (to say the least) to direct against him arguments crucially assuming the legitimacy of such appeals *without* previously answering his many objections of principle to arguments of this sort. It may be noted, incidentally, that Davis makes no reference to any of the numerous independent discussions defending or attacking Davidson's views of act individuation, thus making it impossible to judge exactly where he stands on the many relevant issues raised by these discussions. And it should also be noted that, perhaps most surprisingly given his ostensible topic, he fails to take notice of a study by Robinson (1974) which for the first time attempted to use speech act theoretic materials in testing both Davidsonian and anti-Davidsonian views of act individuation, and has already argued on independent grounds that, at least in their simplest forms, both types of view have difficulties in handling these materials. It seems to me, then, that Davis's paper not only fails to develop a convincing original argument against Davidson's views, but does not even succeed in acquainting its readers with the variety of complex issues that should be addressed by anyone seriously attempting either to oppose or to support those views.

The paper by Tomoyuki Yamada ("An ascription-based theory of illocutionary acts") is a preliminary attempt at developing, on the basis of Devlin's (1991) formalisation of situation theory, a notation capable of yielding definitions of illocutionary acts in terms of their context-changing properties, and, most importantly, of capturing identities of content across illocutionary act types *without* assuming—as is standardly assumed—that the contents to which illocutionary forces get attached are necessarily truth-evaluable entities. Though preliminary in character, the formal proposal is convincingly motivated and clearly implemented, thus making Yamada's paper a worthwhile contribution both to situation theory, where questions of illocutionary force are not systematically addressed, and to speech act theory, where the context-changing nature of illocutionary acts, and so their character as *acts*, is often ignored, while the characterisation of their contents standardly takes the facile route of using truth-theoretic notions in contexts where their relevance, or even their intelligibility, is questionable.

The paper by Bernard Moulin and Daniel Rousseau ("An approach for [sic] modelling and simulating conversations"), is not in any specific sense about speech acts, but its most surprising feature is that, in spite of its title, it is not in any clear sense about conversation either. What the authors are primarily offering is a series of speculations about the proper design of a mechanical device that

would enable the operations executed by two or more robots to be coordinated, and would thus provide, in the authors' view, a simulation of human conversation. In a manner familiar from some other works in artificial intelligence, the authors never explain exactly *why* the coordinated execution of these operations, assuming it to be mechanically possible, should be deemed to be a simulation of a human conversation. Indeed, they do not even seem overly concerned about the fact, revealed in a footnote (p. 292) but studiously obscured in the text, that their robots, though allegedly designed to conduct 'conversations', are *not* supposed to be either language generation or language recognition devices—and, as if that was not enough for doubting that they have a clear conception of what they are trying to model, they describe the phenomena they are interested in modelling as activities where agents "propose [to each other] mental states" and "negotiate about the mental states they propose" (p. 184), without appearing to realise the multitude of category mistakes that this description involves. The authors' underlying idea is, apparently, that, once a device could be constructed that would successfully coordinate the execution of one or more tasks by two or more robots, it would become true *by definition* that the robots are having a 'conversation' of some sort. But if that was their idea, they should make its purely stipulative character explicit rather than pretend that they have offered a simulation of human conversation just because they have elaborated a complex terminological ritual where the various components of their collection of mechanical devices are given names that might directly or indirectly recall real conversational episodes. Notice, for example, that not only are the robots themselves called "locutor [sic] agents"—in spite of the fact that they are not supposed to either generate or recognise anything linguistic—, but that each of the various *parts* of the device that would coordinate their performances is supposed to be a separate 'agent' by itself: The part of the device that would record the robots' past reactions to instructions is called "the conversational agent", the part of the device that would encode data about the robots' environment is called "the environmental agent", the part of the device that would determine whether the robots associate the same or different outputs to a given input is called "the discourse manager", and so on. People interested in speculative artificial intelligence may be fascinated by this kind of parody, but I doubt that the paper would have anything to offer to prospective readers of the V&K volume, assuming that none of these readers would be so gullible as to accept without argument that a mechanism of successful robotic coordination is necessarily a proper model of human conversation. (Prospective readers should also be warned that, though the paper purports to be written in English, its attempted simulation of English does not always produce acceptable results. For example, instead of 'relations like acceptance, denial,

etc.', the authors write, "relations like acceptation [sic], denial, etc." (p. 187); and instead of (presumably) 'one piece of evidence for that claim is [x]', they write, "an [sic] evidence of [sic] that claim is [x]" (p. 184).)

Although its title would appear to promise grander things, the paper by Susumu Kubo ("Illocutionary morphology and speech acts") is about a *single* verbal affix of Japanese (*teyaru*), which the author proposes to call an "illocutionary affix", in the manifest hope of thereby establishing a new field of research called "illocutionary morphology". The author's highly experimental English may make some readers unable to read through the paper (for example, in introducing it, Kubo writes, "This paper *will be* written on the basis of the philosophical ideas of speech act theory" (p. 211; italics added), a statement that may make some readers doubtful as to whether the paper they appear to be reading has or has not been written yet). However, even readers persistent enough to read through the paper will not, as far as I can predict, either acquire a clear understanding of what *teyaru* is supposed to mean or detect a clear reason for calling it an "illocutionary affix". Kubo's first attempt at explaining what *teyaru* means is the following (p. 210): "*Teyaru* serves to give the hearer the benefit of the act represented in the propositional content by revising and extending the previous analysis given in Kubo (1993)". This explanation should obviously be discarded on *a priori* grounds, since it implies that the meaning of *teyaru* contains references to Kubo's publications. Kubo's subsequent attempts at explaining what *teyaru* means vacillate between the claim that it is an indicator of speaker status (for example, "*teyaru* is used in (...) contexts (...) in which the speaker invokes a position of the [sic] authority over the hearer" (p. 214)), and the claim that it is an indicator of illocutionary force (for example, "*teyaru* names either commissive or declarative illocutionary forces" (p. 211)). Since Kubo treats these explanations as equivalent, a reasonable hypothesis is that he is actually *confusing* status indicators with force indicators. That hypothesis is fully confirmed by a footnote, where Kubo writes, "*teyaru* has the illocutionary meaning, 'the speaker's authority over the hearer'" (p. 294)—without, presumably, realising that, if the expression "the speaker's authority over the hearer" names anything at all, it certainly does not name an illocutionary force of any kind, and so cannot be supposed to have an "illocutionary meaning" in any known sense of that term. Keeping the notions of status indicators and force indicators clearly distinct, one might, nevertheless, examine whether the data actually presented in the paper favour an interpretation of *teyaru* as a force indicator, in which case Kubo's decision to call the affix an "illocutionary affix" would be (in spite of his explanations) vindicated, or, on the contrary, favour an interpretation of *teyaru* as a status indicator, in which case the decision to call it an "illocutionary affix" would be blatantly unjustified.

Upon examination, it turns out that every single instance of the affix's use cited in the paper strongly suggests that *teyaru* is a status indicator (a hardly surprising result, perhaps, given the omnipresence of status distinctions in Japanese morphology) rather than a force indicator, even though Kubo, always unable to maintain a clear conceptual distinction between status and force indicators, persists in misdescribing it as a force indicator. For example, the explicit performative utterances *Ayama-t-teyaru*, *Kansha-shi-teyaru* and *Iwa-t-teyaru* are respectively glossed (p. 214) as 'In my capacity as your superior, I hereby apologise to you', 'In my capacity as your superior, I hereby thank you' and 'In my capacity as your superior, I hereby congratulate you'. Assuming that they are even approximately correct, these glosses make it perfectly clear that, in each case, it is the performative verbal stem rather than the *teyaru* affix that determines the utterance's illocutionary force, whereas *teyaru* merely invokes the speaker's presumed 'superior' status relative to the hearer. In spite of this, Kubo describes the above examples as confirmatory of the following 'generalisation': "The illocutionary force of an utterance with [the] illocutionary affix *teyaru* is not that of the matrix verb of the utterance, but that of the illocutionary affix" (p. 216). This statement implies that, contrary to what Kubo's own glosses show, the above mentioned performative utterances do *not* realise the acts of apologising, thanking and congratulating named by their verbal stems, but rather the hitherto unknown "illocutionary act" named by *teyaru*, which, as already noted, Kubo glosses as "the speaker's authority over the hearer". Since, however, "the speaker's authority over the hearer" is not the name of any illocutionary act whatsoever, and since the illocutionary acts of apologising, thanking and congratulating, which are named in the relevant verbal stems, are certainly performable by the utterances Kubo cites, the only option consistent with Kubo's actual data is that of describing *teyaru* as a status indicator and *not* as a force indicator. And in that case, of course, the whole idea on which the paper rests, namely, that *teyaru* is a force indicator whose existence would justify the establishment of a new field of research called "illocutionary morphology", should obviously be rejected. Exactly where Kubo's confused identification of status indicators with force indicators derives from is difficult to tell, but some remarks (p. 214, p. 215) suggest that it is related to the fact that he has read in Searle and Vanderveken (1985) that there is something called a "mode of achievement of an illocutionary force", which sometimes concerns authority relations between speakers and hearers as *preconditions* on the performance of certain illocutionary acts. Obviously, however, Searle and Vanderveken (1985) are not responsible for Kubo's confusions, since, just as a *mode of catching* a fly is not a fly, so a *mode of achievement* of an illocutionary

force is not an illocutionary force; consequently, the fact that a speaker is or purports to be in a position of authority when performing a given illocutionary act does not, contrary to what Kubo imagines, make positions of authority *identical* to illocutionary acts.

Although one wouldn't suspect it from its title, the paper by Masaaki Yamanashi ("Speech act constructions, illocutionary forces and conventionality") is mainly about compatibility constraints on quoting verbs and quoted utterance types in contexts of direct quotation. The paper assumes no theory of quotation whatsoever, and so it is unobvious exactly what the significance of its claims could be supposed to be, even if they were true. Besides, to the extent that the claims are both clear and non-trivial, they do not appear to be true. The one among them that is least unclear and arguably non-trivial can be formulated as follows: If an utterance is an indirect but *conventionalised* way of performing a given illocutionary act, then that utterance can be introduced in a context of direct quotation by a verb naming the illocutionary act in question; whereas if an utterance is an indirect but *non-conventionalised* way of performing a given illocutionary act, then it cannot be introduced in a context of direct quotation by a verb naming the illocutionary act in question. For example, *I want you to leave this room* and *Your guests are waiting for you* can both be indirect requests that the hearer should leave a certain room, but only the former, which is conventionalised, and not the latter, which is not conventionalised, may be introduced by a verb such as *request* in a context of direct quotation: *"I want you to leave this room", he requested* is acceptable, whereas *"Your guests are waiting for you", he requested* is unacceptable. The problem, however, is that the author's generalisation (which the paper would anyway leave entirely unexplained, even if it obtained) does not invariably obtain. On the one hand, an indirect *and conventionalised* request of a hearer's leave can be effected both by *I must now ask you to leave* and by *I am now going to ask you to leave*; contrary, however, to what the author's generalisation entails, neither of these utterances can be acceptably introduced by the relevant illocutionary verb in a context of direct quotation: both *"I must now ask you to leave", he asked me* and *"I am now going to ask you to leave", he asked me* are unacceptable. On the other hand, there can be indirect and *non-conventionalised* requests of a hearer's leave—for example, *I don't want your guests to wait for you any longer*—that, contrary to what the author's generalisation entails, *can* be acceptably introduced by relevant illocutionary verbs in contexts of direct quotation—for example: *Mary asked John if there was anything he wanted her to do. "I don't want your guests to wait for you any longer", he requested*. The upshot is, then, that the possibility or impossibility of acceptably

introducing a direct quotation of a given utterance by an illocutionary verb cannot be systematically correlated with the conventionalised or non-conventionalised character the utterance's ability to indirectly implement illocutionary acts named by that illocutionary verb.

The paper by Jacques Moeschler ("Speech act theory and the analysis of conversations") is a highly condensed description of the itinerary of a Geneva-style conversation analyst who, it seems, was initially fairly confident about the utility of speech act notions in constructing a model of conversational structure, has subsequently become rather sceptical about the possibility, or even the intelligibility, of constructing such a model—partly, though not exclusively, under the influence of Searle's (1986b) conjecture that conversations do not *have* an intrinsic structure to be modelled, whether in speech act theoretic or in non-speech act theoretic terms—, and has finally decided to place his bets on relevance theory (Sperber and Wilson 1986), not in the hope of modelling an intrinsic structure that conversations probably don't have but in the hope of showing that conversational structure is a purely epiphenomenal outcome of speakers' and hearers' attempts to interpret each other's vocalisations in accordance with the principle of relevance. This might have been an interesting itinerary for Moeschler to have followed, but those curious about it would be well advised to consult his various publications that the paper summarises rather than the paper itself. The accounts of the itinerary's abrupt changes of direction are far too brief; the terminology is often idiosyncratic (for example, "pragmatic theory" and "relevance theory" are used as synonyms); the formulation of central claims is often insufficiently precise (for example, the author's currently favoured position on the relation between sequencing phenomena and interpretation phenomena is supposed to be conveyed both by the claim that "the sequencing problem is part of the interpretation problem" (p. 250) and by the claim that "the sequencing problem is equivalent to the interpretation problem" (p. 258), although these two claims are not equivalent, on a normal understanding of their terms); and, last but not least, the identity of the language in which the paper is written is often unclear (for example, there is no English verb 'explicit' as there is a French verb 'expliciter', but the paper, though apparently written in English, contains word strings like the following: "The first purpose of this paper is to explicit the divergence between [x] and [y] about [z]" (p. 239)).

Marc Dominicy and Nathalie Franken begin their paper ("Speech acts and relevance theory") by warning that it will not deliver what its title would appear to promise—namely, a comprehensive examination of relevance-theoretic claims about the analysis of speech acts—, and they indeed appear to be unaware of an important study by Bird (1994) which, having taken the first steps towards such

an examination, has reached results that, if taken into account, would help placing their investigations in a proper perspective. Nevertheless, the more restricted goal that their paper primarily pursues—that of evaluating the relevance-theoretic analysis of the imperative mood—is expertly and insightfully accomplished. Since the relevance-theoretic analysis in question was originally presented as an attempt to overcome allegedly insuperable problems that a speech act theoretic conception of the imperative mood would face, the authors' examination has two sub-goals: That of showing that a particular version of speech act theory can in fact account, if appropriately deployed, for most of the uses of imperatives that, according to relevance theorists, speech act theories would be unable to uniformly handle; and that of showing that the alternative account of the imperative mood that relevance theorists have themselves proposed fails, upon closer examination, to give a satisfactory account of most of the uses of imperatives they have set out to cover. The observations and arguments that the authors offer while pursuing each of these sub-goals are unfailingly interesting. The paper, therefore, should supply appropriate material for discussion both to relevance theorists and to speech act theorists preoccupied with the analysis of mood—although the latter might happen to know, having read Harnish (1994), that a version of speech act theory more parsimonious than the one that the authors use has already offered a unifying account of imperatives that sidesteps most of the relevance theoretic objections.

I will conclude with Daniel Vanderveken's paper, which opens the V&K collection. Upon reading its title ("Universal grammar and speech act theory"), readers might form the expectation that its forty pages will contain some kind of discussion of the question whether speech act theoretic notions could or should be incorporated into contemporary Chomskyan syntactic theory. In fact, the paper offers nothing of the sort, and is entirely unrelated to the few recent works in the Chomskyan tradition (see, for example, Culicover 1992; Rizzi 1997; Han 2000; Krifka 2001; Haegeman 2003) that explicitly rely on illocutionary notions in describing grammatical phenomena. Rather, the author's two main goals are (a) to give an informal exposition of some ideas of the system of illocutionary logic that was officially presented in Searle and Vanderveken (1985) as a formalisation of some aspects of Searle's theory of speech acts, and (b) to suggest how some elements of that system could be used in reformulating Grice's theory of conversational implicatures. A parallel goal appears to be the expression of the author's recently formed opinion that, though originally unrelated to any sort of Kantianism, illocutionary logic might be viewed as a kind of vindication of Kantian transcendentalism; that opinion, however, cannot be seriously considered in the

form in which it is expressed, not only because its exposition completely and unaccountably ignores the work of Karl-Otto Apel, which is the most sustained contemporary attempt at a transcendentalist interpretation of speech act theoretic ideas (see, for example, the papers recently collected in Apel 1994), but mainly because, to the minimal extent to which it is defended by the author, it reveals quite fundamental misunderstandings of basic Kantian notions. For example, although Vanderveken claims that "illocutionary logic is transcendental in the sense of Kant" (p. 36), his main way of substantiating that claim consists in attaching the label "transcendent" (p. 26, p. 35, p. 36, p. 37, p. 45, p. 46) to every single thesis of illocutionary logic that he repeats, without ever realising something that even beginning students of Kant know, namely, that the terms "transcendent" and "transcendental" have radically different meanings in Kant's system and that, in particular, if something is "transcendent", then, according to Kant, not only could it not possibly be "transcendental", but would lie entirely outside the limits of possible thought. Vanderveken's contributions to the topics listed under (a) and (b) above will be examined in that order.

Considering that, in the nearly twenty years since the official presentation of illocutionary logic, Vanderveken has been its only promoter, the paper's attempt to provide a summary exposition of some of the logic's main ideas might be expected to be, if not particularly exciting, at least reasonably accurate. It turns out, however, that it is not even that. For one thing, the official presentation, unlike Vanderveken's summary, does not attempt to create the illusion that illocutionary logic is a contribution to Kantian philosophy (it does not, for example, contain the obviously untrue claim that the axiomatically introduced 'five illocutionary points' have been established through some kind of "transcendental deduction" (p. 61)), nor does it attempt to create the illusion that illocutionary logic is a contribution to generative grammar (it does not, for example, indulge in the gratuitous activity of baptising illocutionary points "material linguistic universals" (p. 32) and operations on illocutionary forces "formal linguistic universals" (p. 35))—notice, incidentally, that if, as the above characterisations jointly imply, illocutionary points were both 'material linguistic universals' and 'transcendentally deducible' entities, it would follow that one could discover material linguistic universals just by means of transcendental deductions, which is, I believe, as grotesque a claim as one could manage to make either about material linguistic universals or about transcendental deductions.) For another thing, the official presentation of illocutionary logic, unlike Vanderveken's summary, generally avoids the practice of presenting intelligible theses as if their acceptance required the acceptance of unintelligible premises. Limitations of space allow mention of just two examples. At one point Vanderveken claims that, according

to illocutionary logic, the "imperative sentence *Please, help me!* truth condition-
ally entails the declarative sentence *You can help me*" (p. 48); illocutionary logic,
however, does not officially endorse the incoherent idea that imperative sen-
tences, even though they do not have truth conditions, *truth conditionally* entail
certain other sentences that do have truth conditions; what it endorses is the un-
surprising claim that if certain *statements* about imperatives are true, then certain
other *statements* about imperatives must also be true. For example, if it is true
that, in order to comply with the imperative *Please, help me!*, the hearer must *help*
the speaker, then it also has to be true that, in order to comply with the same
imperative, the hearer must *have the ability to help* the speaker—simply because
it is logically impossible to provide help without having the ability to provide
help. And although what illocutionary logic would actually claim about this and
similar examples is infinitely less innovative than Vanderveken's summary sug-
gests, it is the only thing that it *could* claim without lapsing into the kind of inco-
herence that the summary exemplifies. At another point Vanderveken claims that
illocutionary logic can "predict and explain why we are all able to infer from the
premise *Please, give me a glass of red or white wine!* the conclusion *Please, give
me a glass of wine!*" (p. 50); illocutionary logic, however, does not officially en-
dorse the questionably intelligible idea that, given an imperative sentence ut-
tered by a speaker, we can 'infer' that the same speaker has uttered certain other
imperative sentences that he has not in fact uttered; what it endorses is the un-
surprising claim that, given a *statement* describing what a speaker has done in
uttering an imperative sentence, certain other *statements* necessarily follow. For
example, if it is true that, in uttering a certain imperative sentence, a speaker has
attempted to obtain wine that is either red or white, then it must also be true that
the same speaker, in uttering the *same* imperative sentence, has *attempted to ob-
tain wine*—simply because it is logically impossible to attempt to obtain wine that
is either red or white without at the same time attempting to obtain wine. And
although, again, the realisation that illocutionary logic ultimately deals with in-
ferential relations between certain kinds of *statements* makes it far less original
than Vanderveken's summary suggests, it fortunately makes it far less incredible
than the same summary implies. Finally, the official presentation of illocutionary
logic is written in readily interpretable English, whereas Vanderveken's summary
is written in a kind of English whose interpretation requires the possession of
some unnatural abilities (for example, the ability to read "empiric" (p. 61) and
understand 'empirical', to read "disambiguous" (p. 30) and understand 'unam-
biguous', to read "insatisfaction" (p. 60) and understand 'non-satisfaction', to
read "I use to be" (p. 52) and understand 'I used to be', etc.) It seems to me, then,

that the official presentation of illocutionary logic is a much better source to consult in order to find out just what the theses of that logic are (and also, incidentally, just how unoriginal many of these theses are, when they are true) than the inaccurate and misleading summary supplied by Vanderveken's contribution.

Coming, finally, to Vanderveken's suggestions as to how (his version of) illocutionary logic could lead to an improved reformulation of Grice's theory of conversational implicatures (see Grice 1975 and the relevant papers reprinted in Grice 1989), it seems to me that the suggestions are seriously misguided both in what they assume and in what they propose. Vanderveken accepts that Grice's theory is unproblematic as far as implicatures associated with *assertive* utterances are concerned, but is worried that the theory is not equipped to deal with implicatures associated with *non-assertive* utterances, and so he proposes generalised reformulations of Gricean conversational maxims in illocutionary terms, believing that these reformulations will make the maxims capable of accounting not only for implicatures associated with assertive utterances but also for implicatures associated with non-assertive (in particular, directive) utterances. To cite the one reformulation that is sketched in some detail, Grice's maxim of quality should, according to Vanderveken, be reformulated as, "Let the illocutionary act that you mean to perform be felicitous in the context of your utterance" (p. 53), where an illocutionary act is defined as felicitous in a context of utterance just in case it is "successful, non-defective and satisfied" (p. 53) in that context. It is Vanderveken's belief that, applied to the special case of assertive utterances, the generalised formulation will engender exactly what Grice's original quality maxim was demanding, and will thus contribute in the well-known Gricean way to the explanation of assertively induced implicatures; but the generalised formulation of the maxim will *also* be applicable, Vanderveken contends, to all kinds of non-assertive utterances (in particular, to directive utterances), and it will thus make possible the construction of analogous explanations of non-assertively induced implicatures as well. Now, Vanderveken's entirely undefended assumption that, as far as assertively induced implicatures are concerned, the Gricean explanations are unproblematic is, after so many years of critical attention devoted to the Gricean approach, difficult to believe, and can only be responded to by referring him to, say, Wayne Davis's recent book (Davis 1998), where many of the severe conceptual and empirical problems that the Gricean approach faces are ably analysed. But even assuming that the Gricean approach *is* valid for assertively induced implicatures, Vanderveken's claim that his proposed generalisation of the Gricean maxims can successfully account for both assertively and non-assertively induced implicatures is demonstrably mistaken.

Recall that Vanderveken's generalised maxim of quality requires that illocutionary acts should be felicitous in the context of utterance, and that the notion of 'felicity' involved is explicated by specifying that the illocutionary acts in question should be successful, non-defective *and satisfied*. Now, according to illocutionary logic, an assertive illocution is satisfied when it is *true*, whereas a directive illocution is satisfied when it is *complied with* (that is, when it is obeyed if it is an order, when it is followed if it is a suggestion, when it is granted if it is a request, etc.) Consequently, applied to assertive utterances, the generalised maxim of quality entails that the assertive illocutions expressed by these utterances should be *true*—which, Vanderveken supposes, is exactly what, according to Grice, hearers would have to assume in order to be in a position to calculate many assertively induced implicatures. Grice, however, was not claiming that, in order to calculate these assertively induced implicatures, the hearer should take the speaker's assertions to be true, but merely that he should take these assertions to *be believed* by the speaker to be true. And Grice was, of course, quite right in doing so, since, if he were to adopt Vanderveken's requirement, he would be quite unable to derive a large number of assertively induced implicatures that his own requirements do allow him to derive. It is clear, for example, that a speaker who answers the question, *Where is Mary?*, by the assertion, *Mary's car is parked outside Helen's house*, can successfully *implicate* that Mary is probably inside Helen's house, *even if* his hearers happen to *know* that his assertion is factually incorrect (even if they know, for example, that the car to which the speaker purports to be referring by the phrase *Mary's car* is, unbeknownst to the speaker, not Mary's but someone else's car). Grice would have no difficulty in deriving the implicature in that context, since all that, according to him, the hearers would need to assume for their calculations is, on the one hand, that the speaker *believes*, no matter how mistakenly, that it is Mary's car that is parked outside Helen's house, and, on the other, that he is trying to be *cooperative* in his answers; but Vanderveken could not possibly derive the implicature in the same context, since his generalised maxim would require the hearers not only to take the speaker *to be sincere* in his assertions but also to take him *not to be mistaken* in his assertions (for it is only by taking the speaker not to be mistaken, and not merely by taking him to be sincere, that they could take his assertions to be *true*, and not merely *believed* by him to be true); and since a huge number of implicatures are communicated through assertions that, though known to be sincerely made, may also be known to be factually incorrect, there is a huge number of assertively induced implicatures that can easily be derived by Grice's original quality maxim but can-

not possibly be derived by Vanderveken's generalised quality maxim. So, contrary to what Vanderveken supposes, his generalised maxim does not even succeed in recapturing Grice's account of assertively induced implicatures.

The situation is just as bad, however, when the generalised quality maxim, in the course of Vanderveken's attempt to explain something that Grice never attempted to explain, namely how *non*-assertively induced implicatures are calculated, is applied to directive utterances. For, applied to directive utterances, the generalised quality maxim requires that the directive illocutions expressed by such utterances should be felicitous; and this, given Vanderveken's definition of felicity, entails that the directive illocutions in question should be satisfied (that is, *complied with*). Assuming, then, that, as Vanderveken explicitly states, the fundamental source of these non-assertively induced implicatures is the hearer's hypothesis that each one of the conversational maxims (and so, the generalised quality maxim) has been respected by the directive illocutions communicating the implicatures, it follows that no such implicatures could be derived from any directive utterance unless the hearer was taking the directive illocution expressed by such an utterance to have respected the generalised quality maxim by having been felicitous (and so, by having been *complied with*). But this amounts to saying that the addressee of a directive utterance cannot derive an implicature communicated by that directive utterance unless he takes himself to *have complied with* that directive utterance. And that consequence is so obviously absurd that no theory of implicatures that generates it has any chance of being accepted. Suppose, for example, that, in response to a speaker A's assertive utterance, *I wasn't able to find Mary in Paris*, a speaker B, who had previously asked A to find Mary and who still wants A to find Mary, produces the directive utterance *Go to Rome!*, thereby implicating that A may be able to find Mary in Rome. On Vanderveken's account, A will not be able to calculate what B implicates unless he takes B's directive utterance to have respected the quality maxim, and so to have been felicitous. But, on Vanderveken's account of felicity, A cannot take B's directive utterance to have been felicitous unless he takes it to have been *satisfied* (that is, *complied with*). And since B's directive utterance cannot be supposed to have been complied with unless A is supposed to have actually gone to Rome, it follows that, according to Vanderveken, A cannot as much as *understand* what B implicates by saying *Go to Rome!* unless he takes himself to have actually gone to Rome. But that consequence is clearly absurd: In order to understand that, when uttered in response to *I wasn't able to find Mary in Paris*, the imperative *Go to Rome!* can successfully implicate that its addressee may be able to find Mary in Rome, it is certainly not necessary to assume that either its addressee or anyone else *has ever gone or will ever go to Rome*. And any theory that, like

Vanderveken's, is constrained to deny this simple fact, can hardly be considered, let alone accepted, as a defensible account of non-assertively induced implicatures. Adding to this that, for reasons already explained, Vanderveken's theory cannot supply a defensible account of even assertively induced implicatures, it is difficult to avoid the conclusion that Vanderveken's attempted employment of illocutionary logic as a basis for reconstructing Grice's theory of conversational implicatures is an unmitigated failure.

It will come as no surprise, given all the above remarks, that my opinion on the V&K volume is very far from being positive: With the exception of just two papers (the one by Yamada and the one by Dominicy and Franken), the volume makes no appreciable contribution to the advancement of research on any speech act theoretic topic, and its extreme lack of editorial care about linguistic matters risks embarrassing even the least demanding among its readers (though, obviously, it did not succeed in embarrassing its publishers). Overall, then, the V&K volume is a sad but clear example of the low standards of argumentation, and of the even lower standards of publication, that are sometimes operative in contemporary pragmatics.

Chapter 19
Searle's derivation of promissory obligation

1 Introduction

In "How to derive 'ought' from 'is'" (Searle 1964), perhaps the most famous among his early articles, John Searle set out to show that what is sometimes called "the naturalistic fallacy"—the fallacy that is allegedly committed by those who affirm that it is possible to deduce evaluative conclusions from wholly non-evaluative ('descriptive') premises—is not at all a fallacy, contrary to what David Hume and the long list of philosophers who follow Hume on this matter have supposed. The reason that it is not a fallacy, Searle claimed, is that there are certain kinds of clearly evaluative statements (specifically, statements about what a particular person *ought* to do or *has the obligation* to do) that *are* deducible from certain sets of wholly non-evaluative ('descriptive') statements (specifically, from sets of statements that include a statement about what that person has *promised* to do). That being so, Searle contended, not only is there no such thing as "the naturalistic fallacy" but those claiming that there is commit themselves to the denial of the validity of a series of logically impeccable inferences (and so become victims of what might be called "the naturalistic fallacy fallacy").

Searle presented an improved version of his proposed derivation of "ought" from "is" (in effect, of statements about obligations from statements about promises) in the final chapter of his first major book, *Speech Act*s (Searle 1969), arguing that all criticisms of the derivation that in the meantime had been produced have been unsuccessful, and claiming that the derivation is one among the many philosophically interesting by-products of the general account of meaning and speech acts that that book proposes. In outline, Searle's improved version of his derivation proceeds as follows. From a *descriptive* statement to the effect that a speaker X has uttered, at time *t*, a sentence of the from "I (hereby) promise to *p*", where *p* refers to a future action of X, we can infer, Searle says, the equally *descriptive* statement that X has promised, at *t*, to *p*, simply by making the empirical assumption that X's utterance satisfies the conditions that the theory of speech acts specifies as necessary and sufficient for a speaker's utterance to constitute a promise on his part to perform a future action. But once these two purely *descriptive* statements are in place, Searle contends, they logically necessitate, in conjunction with three analytic (and so, non-evaluative) statements, a series of clearly *evaluative* statements, and so contradict the thesis this is impossible to

logically deduce evaluative conclusions from non-evaluative premises. Specifically: (a) in conjunction with the analytically true statement that *whoever promises, at a given time, to do something, undertakes, at that time, the obligation to do it*, the statement that X promised, at *t*, to *p* entails the statement that X undertook, at *t*, the obligation to *p* (e.g., the statement that Tom promised, at *t*, to give Mary five dollars entails the statement that Tom undertook, at *t*, the obligation to give Mary five dollars); (b) in conjunction with the analytically true statement that *whoever undertakes, at a given time, the obligation to do something, has, at that time, the obligation to do it*, the statement that X undertook, at *t*, the obligation to *p* entails the statement that X had, at *t*, the obligation to *p* (e.g., the statement that Tom undertook, at *t*, the obligation to give Mary five dollars entails the statement that Tom had, at *t*, the obligation to give Mary five dollars); and finally, (c) in conjunction with the analytically true statement that *whenever one has the obligation to do something, one ought, as regards that obligation, to do it*, the statement that it was true of X at *t* that he had the obligation to *p* entails the statement that it was true of X at *t* that he ought, as regards that obligation, to *p* (e.g., the statement that it was true of Tom at *t* that he had the obligation to give Mary five dollars entails the statement that it was true of Tom at *t* that he ought, as regards that obligation, to give Mary five dollars). But since statements about what a particular person at a particular time *ought* to do, or *has the obligation* to do, are clearly evaluative statements, it follows, Searle concludes, that such statements *can* be logically deduced from appropriate sets of wholly *non-evaluative* statements, contrary to the Humean claim that it is never possible to logically deduce evaluative statements from wholly non-evaluative ones.

Searle's early thesis about the deducibility of statements about obligations from statements about promises deserves attention not only on account of its intrinsic interest, but also on account of the special significance it has acquired in the context of the ambitious philosophical projects that he has launched in *The Construction of Social Reality* (Searle 1995) and in *Rationality in Action* (Searle 2001). One of the main aims of *The Construction of Social Reality* was to explain, compatibly with a naturalistic world-view, how institutional facts are possible. A key element of the proposed explanation is that institutional facts are facts that only exist because groups of agents collectively acknowledge their existence by virtue of accepting constitutive rules of the form "*X* counts as *Y* in context *C*", whereby status functions are assigned to entities. And since, according to Searle, the assignment of a status function to an entity amounts to the creation of a *deontic power* (that is, of a right or of an obligation, or, more generally, of an entitlement or of a requirement) involving that entity (as Searle succinctly puts it

[1997b: 451], "the imposition of a status-function is the imposition of a deontology"), the fundamental kinds of things that, in the domain of institutional reality, agents bring into existence simply by acknowledging their existence are the deontic powers (e.g., the rights or the obligations) that follow from the collective assignment of status functions to entities. But since, according to Searle, deontic powers (as opposed to brute powers) are not determinable on the basis of the physical constitution of the entities to which status functions are assigned, it is often necessary, in order for a deontic power to be brought into existence, to be publicly and conventionally *represented* as being brought into existence. And since the commonest means of *public* and *conventional* representation is language, language plays, according to Searle, a key facilitating role in the creation and, especially, in the proliferation of institutional facts. Now, if institutional facts are, essentially, impositions of deontologies, and if language plays the indicated key facilitating role in the creation and proliferation of such facts, it should not be surprising to find that an obligation, which is a particular kind of deontic power, is brought into existence by a promise, which is a particular kind of linguistic act. Viewed in the context of *The Construction of Social Reality*, therefore, Searle's early derivation of statements about obligations from statements about promises does not merely represent, if correct, an isolated instance of the dependence of an extra-linguistic institutional fact on a linguistic fact, but rather a paradigmatically clear, though by no means unique, instance of the pervasive dependence that, according to Searle, obtains between facts of the former kind and facts of the latter kind.

Searle's early derivation has an even more important connection with another major theme of his recent work, namely, the critique of the classical model of practical reason that he develops in *Rationality in Action*.[1] One of the fundamental defects of the classical model of practical reason, Searle contends, is that, according to it, the only reasons for action that agents could possibly recognise and have are their desires or other states essentially dependent on their desires. And the fundamental move that, according to Searle, is required in order for that defect to be overcome, consists in acknowledging that, among the reasons that can motivate agents to act, there are not only the *desire-dependent reasons* that agents may recognise, but also the *desire-independent reasons* that they may recognise, and especially those desire-independent reasons that they cannot *fail* to recognise because they intentionally *create* those reasons for themselves. But understanding the significance of such desire-independent reasons for action

---

1 See also Searle (1999) for a preliminary statement of that critique, explicitly connected with the topic of Searle's early derivation.

amounts, in Searle's view, to understanding the significance of linguistically generated deontic powers, and in particular the significance of promissory obligations. For, according to Searle, the prototypical instance of an intentionally created desire-independent reason for action is the *obligation* that an agent creates for himself when he *promises* to perform some future action: that obligation is clearly a reason that the agent has to perform the future action; and it is, furthermore, a reason that, on the one hand, holds quite independently of the agent's desires (even if an agent does not *want* to do what he has promised to do, the fact that, through his promise, he has placed himself under the *obligation* to do it, provides him with a *reason* for doing it) and, on the other hand, cannot fail to be recognised by the agent, since it was not dictated by others but was purposively brought into existence by the agent himself through his act of promising. Indeed, not only are the obligations created by acts of promising paradigm cases of intentionally created desire-independent reasons for action, but it is also the case, Searle suggests, that virtually all other linguistic acts create, in ways that are perhaps less obvious but no less certain, many other obligations of a similar sort (e.g., an assertion creates for its author the obligation to provide, if challenged, evidence for the truth of its content, a request creates for its author the obligation not to obstruct the realisation, by its addressee, of the state of affairs that the addressee has been requested to realise, etc.); and since each of these obligations constitutes a reason for action that holds quite independently of the agent's desires, it follows that virtually every normal use of language constitutes a counterexample to the classical model's assumption that the only reasons for action that an agent might have are, or are essentially dependent on, the agent's desires. Obviously, Searle's early thesis that statements about promises entail statements about obligations acquires deep significance in this context, and that makes it understandable why it is frequently invoked by him throughout *Rationality in Action*. For although, as Searle now stresses, promises are not, in his view, unique among speech acts in creating obligations (and therefore desire-independent reasons for action), their conceptual ties with the creation of obligations are so strong and evident that they provide a model on which all linguistically generated obligations (and, therefore, all linguistically generated desire-independent reasons for action) should be analysed.

The main purpose of the present essay is to argue that Searle's thesis that promises necessarily create obligations is open to a family of decisive counterexamples. The reason why these counterexamples have so far remained unnoticed is, in my view, that most of the critical literature that the thesis has provoked (and which I do not propose to review here) does not concern itself with Searle's cen-

tral contention that statements about promises entail statements about obligations (does not address, that is, what I will henceforth call the *deducibility claim* that Searle makes), but rather effectively grants to Searle that statements of the former kind entail statements of the latter kind and only disputes Searle's *separate* claim that statements of the former kind are purely 'descriptive' whereas statements of the latter kind are purely 'evaluative' (it only addresses, that is, what I shall henceforth call the *categorisation claim* that Searle also makes). My counterexamples concern specifically the deducibility claim to which Searle is centrally committed—the claim, that is, that, necessarily, whenever a statement of the form "X promised to *p*" is true, a corresponding statement of the form "X ought/has the obligation to *p*" is also true—and, if valid, they make it strictly speaking unnecessary to further discuss the categorisation claim: if, contrary to Searle's contention, "X promised to *p*" does *not* entail "X ought/has the obligation to *p*", then Searle would presumably not expect any momentous philosophical conclusions to follow *merely* from the assumption that "X promised to *p*" is 'descriptive' whereas "X ought/has the obligation to *p*" is 'evaluative', even if that assumption turned out to be correct. I believe, however, that it is itself doubtful whether Searle would currently be entitled to that last assumption, given some independent theses that he has recently propounded; and I consequently believe that it is doubtful whether he would *now* be entitled to all the conclusions he was originally aiming to draw from his discussion of the "is"–"ought" question, even assuming, contrary to fact, that statements of the form "X promised to *p*" do entail statements of the forms "X ought to *p*" and "X has the obligation to *p*". I will therefore use the present occasion to first offer some critical reflections on the status of categorisation claim, and will then proceed to the primary task of determining the truth or falsity of the deducibility claim.

2 The instability of the categorisation claim

Assume for the sake of argument that statements of the form "X promised to *p*" do entail corresponding statements of the forms "X ought to *p*" and "X has the obligation to *p*". As Searle recognises, in order for the existence of such entailments to imply anything determinate about the "naturalistic fallacy", it should be pre-theoretically quite obvious that statements about promises cannot be anything else than purely 'descriptive' and that statements about obligations cannot be anything else than purely 'evaluative'. And Searle, in all presentations of his derivation, does take both of these things to be pre-theoretically quite obvious—so obvious, in fact, that he spends no time at all *arguing* in their favour. One might

doubt, however, that pre-theoretical obviousness could be reasonably assumed in this area (after all, 'descriptive' and 'evaluative' are not terms like 'polysyllabic' and 'monosyllabic'), and one might accordingly expect that, given different theoretical frameworks within which the labels 'descriptive' and 'evaluative' are being deployed, different decisions as to how these labels should be applied to statements about promises or to statements about obligations would be called for. I suspect that Searle would simply deny that the conditions of application of the labels 'descriptive' and 'evaluative' are theory-sensitive in that sense, and that they lack the degree of stability that his categorisation claim presupposes. I propose to show in this section, however, that such a denial would be unwarranted, since some of Searle's *own* recent theoretical positions commit him to the thesis that statements about promises are no less *evaluative* than statements about obligations, contrary to his explicit early claim that such statements, unlike statement about obligations, are obviously not evaluative.

The source of the instability is Searle's thesis, frequently asserted in *The Construction of Social Reality* and in subsequent works, that *linguistic facts are institutional facts*—a thesis by which Searle do not mean, of course, that the human capacity for producing speech sounds is itself an institutional phenomenon, but rather that the particular systems of symbolising conventions that particular human communities create by exercising that capacity are institutional phenomena. The fact that a member of a human community has produced a sequence of sounds is, Searle admits, not in itself an institutional fact. But the fact that the very same sequence of sounds is a word, or a sentence, in one human community and neither a word nor a sentence in another human community, or the fact that what are phonetically the same words, or the same sentences, in two human communities may have totally different semantic contents in these two communities are, Searle insists, clearly institutional facts—as a passage in *The Construction of Social Reality* (Searle 1995: 99) explains, such terms as "word", "sentence" or "semantic content" do not refer to intrinsic properties of the sound structures that are produced by members of a linguistic community, but rather to "statuses" that have been collectively "imposed" on those sound structures by members of that community. Indeed, not only are linguistic facts, in that sense, institutional facts, but they are, according to Searle, *paradigmatic* institutional facts. Thus, in giving what he takes to be especially clear examples of institutional facts that are created out of brute facts through the imposition of "status functions on entities" in accordance with "the general formula 'X counts as Y in C'", Searle writes:

> I make noises through my mouth. So far, that is a brute fact: there is nothing
> institutional about it. But, as I am a speaker of English addressing other Eng-
> lish speakers, those noises *count as* the utterance of an English sentence;
> they are instances of the formula 'X counts as Y in C'. But now, in an utterance
> of that English sentence, the Y term from the previous level now functions as
> the X term at the next level. The utterance of that English sentence with those
> intentions in that context *counts as*, for example, making a promise. (Searle
> 1998: 128–129; original emphasis)

On Searle's current view, then, both the fact that a sequence of sounds in a *sen-
tence* of a particular human language and the fact that the utterance of that se-
quence of sounds constitutes the performance of a particular *illocutionary act* in
that language, are *institutional* facts holding in the community of humans that
are speakers of that language.

But now recall that, according to Searle, *whenever* an institutional fact is
brought into existence through the assignment of a status-function to an entity
certain *deontic powers* are necessarily created—that is, certain rights and obliga-
tions (or more generally, certain entitlements and requirements) that did not pre-
viously exist come to exist within the human group that collectively acknowl-
edges the fact's existence. If that is so (if, to put it in Searle's own words,
"everything turns out to be deontic" (1995: 109) in the domain of institutional re-
ality), then it follows that the *mere* fact that a sequence of sounds is the utterance
of a *sentence* of a particular language, or the *mere* fact that the utterance of that
sentence is the performance of a particular *illocutionary act* in a particular lan-
guage, are already *deontic* facts of particular kinds, each associated with its pro-
prietary set of entitlements and requirements. But if that is true, and if it is also
true, as Searle is independently committed to holding, that *every* statement of a
deontic fact is an *evaluative* statement, then it follows that the mere statement
that a sequence of sounds is the utterance of a *sentence* of a particular language,
or the mere statement that the utterance of that sentence constitutes the perfor-
mance of a particular *illocutionary act* in that language, are already *evaluative*
statements of particular kinds.

The upshot of all this is, of course, that *every* singular statement figuring in
the derivation of "ought" from "is" that Searle had originally proposed turns out
to be, by his own stipulations, and contrary to his original intentions, ultimately
evaluative. Searle's derivation, it will be recalled, was supposed to show that
from certain purely *non-evaluative* singular statements such as (1) and (2)

(1)   Tom uttered the sentence "I hereby promise to give Mary five dollars".

(2)   Tom promised to give Mary five dollars.

one can validly deduce, through the mediation of appropriate analytic truths, certain deontic—and therefore, according to him, unquestionably *evaluative*—singular statements such as those in (3) and (4):

(3)   Tom has the obligation to give Mary five dollars.
(4)   Tom ought to give Mary five dollars.

But, in view of Searle's current conception of linguistic reality as a particular kind of institutional reality, it now turns out that (1) and (2) are *themselves*, in their own subtle way, no less *evaluative* statements than (3) and (4) are, since (1) and (2) report particular *institutional* facts of the linguistic variety, since every institutional fact is ultimately, according to Searle, a *deontic* fact, and since every statement of a deontic fact is, according to Searle, an *evaluative* statement: As it should be clear by now, a sequence of sounds cannot, for Searle, be Tom's utterance of the *sentence* "I hereby promise to give Mary five dollars" unless Tom is a member of a linguistic population where that sequence of sounds is collectively *accepted* (or *counted*) as the utterance of a sentence; and a sequence of sounds cannot, for Searle, be Tom's *promise* to give Mary five dollars unless Tom is a member of a linguistic population where the utterance of that sequence of sounds is collectively *accepted* (or *counted*) as the making of a promise. But if (1) and (2) report facts that are institutional, and therefore, in their own subtle way, no less deontic than those reported in (3) and (4), then, given Searle's non-negotiable assumption that deontic statements are necessarily evaluative, it follows that (1) and (2) are no less evaluative statements (though, of course, less *blatantly* evaluative statements) than (3) and (4) are. Consequently, even if it should turn out, as Searle claims, that, from (1) and (2), one can validly deduce (3) and (4), it would *not* follow, as Searle *also* wanted to claim, that this shows that certain evaluative statements can be deduced from certain *non*-evaluative ones. In short, Searle's early claim that "uttering certain words in certain conditions just *is* promising, *and the description of these conditions needs no evaluative element*" (Searle 1964: 50; second emphasis added) is simply inconsistent with his current view that the conditions under which a series of sounds constitutes a promise are conditions

whose satisfaction transforms the occurrence of that series of sounds into an institutional fact by imposing on it a particular *deontology*, and are, therefore, as full of 'evaluative elements' as any deontological condition is.[2]

Of course, it would be open to Searle to try to avoid this result by radically modifying his theory of institutional facts in a way that would allow him to withdraw his current commitment to regard statements like (1) and (2) as subtly evaluative, and so to restore his early claim that such statements are definitely not evaluative. But not only do I not see any basis on which such a move could be motivated within Searle's theory as it is now developed, but it is clear to me that, if attempted, it would only reinforce the point with which the discussion in this section has begun, namely, that the question of the applicability of the labels 'evaluative' or 'non-evaluative' to various kinds of statement cannot be answered in the theory-neutral way that Searle was taking to be possible when he originally proposed the derivation of statements about obligations from statements about promises. If that is so, the only sufficiently well-defined issue that remains to be discussed with respect to Searle's derivation does not concern the categorisation claim that it embodies (nor, consequently, the implications or non-implications of that claim on the "naturalistic fallacy"), but rather the deducibility claim that it embodies. It is to the discussion of that claim that I now turn.

3 The falsity of the deducibility claim

In order to show that, contrary to Searle's deducibility claim, statements about promises do not entail statements about obligations, it would be sufficient to show that there exist conditions under which it is possible for a statement of the form "X promised to p" to be true of an individual at a given time but impossible for corresponding statements of the forms "X has the obligation to p" or "X ought to p" to be true of that individual at that time.[3]

Such conditions clearly exist. Suppose that Tom has just heard a story about Othello and Desdemona, but does not know that, unlike himself, Othello and Desdemona are merely creatures of an author's imagination. If, outraged by what the

---

2 It is worth noting that, as the second sentence in the above quotation from Searle makes clear, Searle's original position was not merely that "promise" is not an 'evaluative' word, but that none of *the conditions under which* a series of sounds constitutes a promise involves any evaluative element.

3 I will be assuming with Searle that "ought"-statements have at least one sense in which they clearly ascribe obligations, and that this is the sense of their occurrence in Searle's derivation.

story represents as Othello's mistreatment of Desdemona, Tom says, at *t*, "I hereby promise to kill Othello", or "I hereby promise to protect Desdemona", the fact that neither Othello nor Desdemona exist has no power to make it false that Tom has *promised*, at *t*, to kill Othello and to protect Desdemona. But the fact that neither Othello nor Desdemona exist certainly makes it false that Tom *has the obligation* to kill Othello, or that Tom *has the obligation* to protect Desdemona. An imaginary thing cannot be a thing that one *has the obligation* to kill, or that one *has the obligation* to protect, though it *is* a thing that one can, in one's ignorance, *promise* to kill or *promise* to protect (and, incidentally, it is also a thing that one can, in one's ignorance, *believe* to be worth killing or *believe* to be worth protecting). For that reason, neither a statement like (5) nor a statement like (6) is contradictory:

(5)   You did promise to kill Othello, Tom, but, since Othello doesn't exist, there is, so far, no one that you have the obligation to kill.

(6)   You did promise to protect Desdemona, Tom, but, since Desdemona doesn't exist, there is, so far, no one that you have the obligation to protect.

And the fact that neither a statement like (5) nor a statement (6) is contradictory suffices for showing that, contrary to Searle's deducibility claim, it is *not* the case either that "Tom promised to kill Othello" entails "Tom has the obligation to kill Othello" or that "Tom promised to protect Desdemona" entails "Tom has the obligation to protect Desdemona".

Vacuous names (that is, names without referents, such as "Othello" and "Desdemona") present an obvious problem for Searle's derivation. Since Atlantis is a non-existent city, one can be under no *obligation* to liberate Atlantis; but this does not make it impossible for one to *promise* to liberate Atlantis, if one mistakenly believes both that such a city exists and that it is unjustly occupied. Since Sherlock Holmes is only a fictional detective, one can be under no *obligation* to hire Sherlock Holmes for the investigation of a robbery; but this does not make it impossible for someone to *promise* to the victims of a robbery to hire Sherlock Holmes for the investigation of that robbery, if one mistakenly believes, or simply wants to lead *others* to mistakenly believe, that Sherlock Holmes is a real detective. Beliefs in the existence of non-existent entities may sometimes be sufficient in order for promises involving those entities to be possible, but they are never sufficient in order for obligations involving those entities to be possible.

Vacuous names, however, are not the only, or the commonest, kinds of expression that show Searle's deducibility claim to be untenable. What I shall here

call *vacuous descriptions* (that is, definite descriptions that, as a matter of contingent fact, nothing satisfies) give rise to a far wider, indeed in principle unlimited, range of counterexamples.[4] Suppose that Tom, confusing me with another person, incorrectly believes that I lost my wallet, even though the description "my lost wallet" is vacuous, since, in fact, I never *had* a wallet to lose. Now, if Tom says to me, "I hereby promise to find your lost wallet", I can without contradiction say,

(7)   Tom promised to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else.

What I *cannot* say without contradiction, however, is,

(8)   Tom ought to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else.

If I never had, and so I never lost, a wallet, then it cannot be the case that Tom *ought* to find my lost wallet; but it *can* be the case that Tom, confused as he was, has *promised* me to find my lost wallet. And since, in that circumstance, "Tom promised to find my lost wallet" will be true even though "Tom ought to find my lost wallet" is not true, the former statement cannot entail the latter, contrary to what Searle's deducibility claim requires. Similarly, suppose that Tom, confusing his own exam paper with somebody else's exam paper, and incorrectly believing that his own exam paper contains a spelling mistake, says, "I hereby promise to correct the spelling mistake in my exam paper", even though the description "the spelling mistake in Tom's exam paper" is vacuous, since, in fact, Tom's exam paper contains no spelling mistakes at all. I could then without contradiction say,

(9)   Tom promised to correct the spelling mistake in his exam paper, but he must be confusing his own exam paper with somebody else's, since his exam paper contains no spelling mistakes at all.

What I could *not* say without contradiction, however, is,

---

4 In what follows, I am assuming the 'attributive' reading of definite descriptions, whose existence no one, as far as I know, doubts. I note in passing that Searle belongs to those who believe, in addition, that the 'attributive' reading of descriptions is the only semantically (as opposed to pragmatically) available one. See Searle (1979: 137–161).

(10)   Tom ought to correct the spelling mistake in his exam paper, but he must
         be confusing his own exam paper with somebody else's, since his exam
         paper contains no spelling mistakes at all.

If Tom has made no spelling mistake, then it cannot be the case that Tom *ought*
to correct his spelling mistake, but this does not prevent Tom from incorrectly
believing that he has made a spelling mistake, and so from *promising* to correct it.
And since, in these circumstances, "Tom promised to correct his spelling mis-
take" can be true even though "Tom ought to correct his spelling mistake" cannot
be true, the former of these statements cannot entail the latter, contrary to what
Searle's deducibility claim requires. In short, the existence of vacuous descrip-
tions, no less than the existence of vacuous names, makes the inference from "X
promised to *p*" to "X ought to *p*" invalid, and thus shows Searle's fundamental
claim that promises necessarily create obligations to be untenable. Indeed, since
the vast majority of linguistically possible definite descriptions are such that
merely knowing what they mean does *not* amount to knowing whether or not they
are vacuous, I suspect that the number of promises that have been exchanged in
the course of human history without *any* obligations having thereby been gener-
ated must be staggering.

　　The pertinent generalisation could be formulated as follows. Call "content-
clause" the clause *p* that, in statements of the form "X promised to *p*" and "X
ought/has the obligation to *p*", specifies the content of the promise or of the ob-
ligation that such statements ascribe. Then, if the content-clause *p* contains a
proper name or a definite description and is such that, when it occurs *un*-embed-
ded, it is false whenever that name or that description are vacuous, the effect of
embedding clause *p* in matrix sentences of the forms "X promised to *p*" and "X
ought/has the obligation to *p*" is necessarily the falsity of the latter matrix sen-
tence but not necessarily the falsity of the former matrix sentence. Assuming *one*
of the senses of the distinction between extensionality and non-extensionality
that Searle himself recognises[5] (the sense in which a sentence is extensional with
respect to an item it contains if and only if it entails that the purported referent of
that item exists), this can be re-expressed as follows: if the content-clause of a
promise-ascription or of an obligation-ascription is independently certifiable as
extensional with respect to the names or descriptions it contains, then the prom-
ise-ascription is *non*-extensional with respect to those names or descriptions,
even though the obligation-ascription *is* extensional with respect to those names
or descriptions. In short, the difference between "X promised to *p*[. . .*r*. . .]" and

---

5 See Searle (1983: 181ff; 2004: 174ff).

"X ought/has the obligation to p[. . .r. . .]" , where p is a content-clause and r a name or description inside that content-clause, is that, when p is extensional with respect to r, "X ought/has the obligation to p[. . .r. . .]" is *also* extensional with respect to r, even though "X promised to p[. . .r. . .]" is *not* extensional with respect to r.[6] The claim that promises necessarily create obligations results, then, from the failure to recognise that statements about obligations are extensional in a way in which statements about promises are non-extensional. For it is precisely this difference that makes it possible for statements of the latter kind to be true even though statements of the former kind are false, and it is, in turn, that possibility that prevents statements of the latter kind from entailing statements of the former kind.

It is strictly speaking unnecessary, but it may nevertheless be useful, to point out that the kind of problem exposed above could not be avoided by deploying the sort of strategy that Searle had used in rebutting certain early objections to his derivation that were appealing to the notion of "*prima facie* obligation". Those objections were aiming to dispute Searle's claim that promises entail obligations by relying on the observation that the obligation a person assumes in making a promise may sometimes be in conflict with *other* obligations that that person independently has, and may consequently have to be discharged in order for such a conflict between obligations to be resolved. As Searle has pointed out, that observation, though correct in itself, cannot be construed as an *objection* to his derivation: two obligations cannot be in conflict unless they both *exist*, and an obligation cannot, logically, be discharged unless it is already in existence; so, the fact that an obligation incurred by an act of promising may have to be subsequently discharged in order for a conflict between obligations to be resolved has no tendency to show that it was non-existent *at the time* of promising; and the bringing into existence of an obligation by virtue of an act of promising *at the time* of that act of promising is all that Searle's derivation, properly understood, requires.[7] Notice, however, that this sort of move would be entirely inapplicable to the family of counterexamples considered above. When a vacuous name or a vacuous description figures in the content-clause of a promise, *no obligation whatsoever* is created at the time of promising regarding anything mentioned in the content-clause of the promise (if, for example, I never lost, since I

---

6 I will not examine here the interesting question whether promise-ascriptions or obligation-ascriptions are extensional or non-extensional in senses of that distinction other than the one explained above.

7 Searle's more general misgivings about the standard interpretation of the notion of "*prima facie* obligation" are presented in detail in Searle (1978).

never had, a wallet, and Tom says to me, "I hereby promise to find your lost wallet," no obligation of Tom's to find my non-existent lost wallet is thereby created, no matter what Tom might imagine); but an obligation that has never existed cannot, logically, be discharged, and the question whether the person making such a promise has or does not have any *other* obligations with which his *non-existent* promissory obligation might be in conflict makes no sense. Consequently, the cases considered above cannot be assimilated to cases where a promissory obligation does exist but is *subsequently* discharged because it conflicts with independently assumed obligations; and so, Searle's way of deflating objections to his derivation that rely on situations where conflicts between actually existing obligations are possible is of no use in solving the problem posed by the cases presented above.

Granting the obvious point that, in the cases presented above, no obligations are generated, someone (though, I presume, not Searle) might make a desperate final attempt to save Searle's derivation by claiming that, in the same cases, no *promises* are made either—in other words, that if, for example, Tom says to me, "I hereby promise to find your lost wallet," and it turns out that I never lost, since I never had, a wallet, it not only follows that Tom has not thereby acquired any *obligation*, but also that he has not even made any *promise*. That this is false is evident not only from the already noted fact that, in the described circumstances, Tom's utterance could without contradiction be reported by a statement such as "Tom promised to find my lost wallet, but I never lost, since I never had, a wallet, so he must be confusing me with somebody else", but also from the fact that, in the same circumstances, Tom's utterance would be subject to a variety of forms of appraisal that would be unintelligible *except* on the assumption that it *was* the utterance of a promise. For example, knowing better than Tom whether or not I ever owned a wallet, I might respond to his utterance of "I hereby promise to find your lost wallet" by truly and intelligibly saying to him, "Well, Tom, the promise you just made cannot be fulfilled, since I never had, and so I never lost, a wallet." But that perfectly true and intelligible remark would have to be counted either as just unintelligible or as blatantly false if Tom had *not* made the promise to which I would be referring, and if, in particular, the proposition that I never had, and never lost, a wallet had the power to entail the proposition that Tom did not *promise* to find my lost wallet. What the proposition that I never had, and never lost, a wallet does entail is, precisely, that Tom cannot *fulfil* the promise he has made, not that he cannot *make* the promise he has made. But fulfilling a promise and making it are two quite different things, and the fact that a promise cannot be fulfilled does not imply that it is not a promise, any more than the fact that an

assertion cannot be true implies that it is not an assertion, or the fact that a request cannot be complied with implies that it is not a request. Rather, a promise can be fulfilled *or* unfulfilled only if it is already in existence, just as an assertion can be true or false only if it is already in existence, and a request can be complied or not complied with only if it is already in existence. Indeed, Searle's familiar and fundamental distinction between the *success* of an illocutionary act and the *satisfaction* of its propositional content (see, e.g., Searle 1991a) was designed precisely in order to acknowledge these elementary facts; but none of these elementary facts could be acknowledged if the desperate strategy now under consideration was to be systematically applied, since its application would have the effect of systematically misdescribing what are, in fact, conditions of propositional content satisfaction as if they were conditions of illocutionary act success. It would require claiming, for example, that the existence of a King of Germany is not only a condition on the *fulfilment* of the promise to arrest the King of Germany, but also a condition on the *existence* of the promise to arrest the King of Germany; that it is not only a condition on the *truth* of the assertion that the King of Germany has been arrested, but also a condition on the *existence* of the assertion that the King of Germany has been arrested; that it is not only a condition on *complying with the request* that the King of Germany be arrested, but also a condition on *making the request* that the King of Germany be arrested; and so on. Since no one (and least of all Searle) would presumably be prepared to accept any of these claims, it should readily be granted that the presence of vacuous names or vacuous descriptions in the content-clause of a statement ascribing a promise does *not* necessitate the falsity of that statement. And since, on the other hand, the presence of vacuous names or vacuous descriptions in the content-clause of a corresponding statement ascribing an obligation *does* necessitate the falsity of that statement, there is, as far as I can see, no escape from the conclusion that ascriptions of promises do not *entail* ascriptions of obligations.

4 Conclusion

I have first argued that, even if it were true, as Searle claims, that statements about promises entail statements about obligations, it would not follow, as he also claims, that non-evaluative statements entail evaluative ones (and that, therefore, the "naturalistic fallacy" is not a fallacy), since, as the instability of his own categorisations indicates, the conditions of application of the labels 'evaluative' and 'non-evaluative' to statements about promises or to statements about obligations are far more theory-sensitive than they would have to be if the latter

claim was to follow from the former one. Concentrating, then, on the former claim alone, I have argued that it is not in fact the case that statements about promises entail statements about obligations, since statements about promises can be true whereas corresponding statements about obligations are false when vacuous names or vacuous descriptions occur in the clauses that specify the content of the promises or of the obligations that such statements ascribe—since, in other words, statements about obligations are extensional in a way in which statements about promises are non-extensional with respect to vacuous names or vacuous descriptions appearing in their respective content-clauses.

As indicated in the introductory section, the claim that statements of the form "X promised to p" entail corresponding statements of the forms "X ought to p" and "X has the obligation to p", apart from its early involvement in Searle's discussion of the "naturalistic fallacy", plays a key role in the accounts of sociality and rationality recently developed in Searle's *The Construction of Social Reality* and *Rationality in Action*. In arguing that Searle's claim cannot, at least as formulated, be sustained, I do not, of course, mean to imply that the various innovative theses put forward in these works should simply be discarded. On the contrary, I believe that they deserve sustained examination, and regard the present discussion as a preliminary step in seeking better ways of articulating their content and implications. If one abandons the dogmatic policy of accepting as worthy of one's theoretical attention only those inferential relations that can be *deductively* validated, one will have no difficulty in acknowledging the obvious fact that people do have the tendency to *defeasibly* infer "X ought to p" or "X has the obligation to p" from "X promised to p", in a way in which, for example, they do not have the tendency to infer, even defeasibly, "X ought to p" or "X has the obligation to p" from "X expressed the wish to p". The problem, then, is to explain, in a naturalistic spirit, what it is that warrants inferences of the former sort without warranting inferences of the latter sort, *even though* neither the former nor the latter conform to deductive canons. And it may well be that, once this naturalistic explanation is at hand, it can be made to cohere with many of Searle's ideas about sociality and rationality. For, as Searle would probably agree, sociality and rationality are not entirely, or even largely, deductive affairs.

Chapter 20
Searle's *Making the Social World*

John Searle's book, *Making the Social World* (Searle 2010), will be useful to readers familiar with his work in the philosophy of language and the philosophy of mind, but unacquainted with, and curious to learn about, the 'philosophy of society' that he has been busy building since the mid-nineties. Such readers are offered a lengthy exposition (Chapters 1, 3, 5) of an updated version of the account of institutional facts that was the main theme of Searle's earlier book, *The Construction of Social Reality* (Searle 1995), as well as shorter discussions (mostly drawing on material already presented in two other books, Searle 2001 and Searle 2007) of what Searle perceives as the implications of his account of institutions on issues pertaining to rational action, free will, political power, and human rights (Chapters 6, 7, 8). The book will also be useful to readers who have developed an interest in Searle's account of institutional reality while lacking sufficient exposure to his philosophies of mind and language, since it includes brief overviews (Chapters 2, 4) of his extensive work in these fields, which he presents as providing the foundations of his account of society. Readers already familiar with Searle's major works on mind, language, and society will probably be mainly interested in considering whether the account of institutional facts that he currently adopts differs significantly from the one he had originally proposed, and, if so, whether it places him in a better position than before to attain his stated goals.

Common to Searle's old and new accounts is a conception of institutional facts according to which such a fact (a) cannot exist unless a community collectively accepts it as existing; (b) requires the assignment to an entity of a "status function" (that is, of a function that an entity can only have by virtue of collective recognition, and not merely by virtue of whatever properties it might have prior to such recognition); and (c) characteristically generates, once in existence, "deontic powers" (in particular, rights and obligations) within the community whose behaviour brings it into existence.

One difference between Searle's old and new accounts is that the generation of "deontic powers" is now taken to be a strictly universal consequence, and not merely, as was previously the case, a *nearly* universal consequence, of an institutional fact's creation (Searle 2010: 24). Another difference between the two accounts is that, whereas Searle was previously affirming that no form of collective intentionality, and in particular no form of collective recognition, is reductively

analysable as a logical construction out of instances of individual intentionality, or of individual recognition, he now admits (Searle 2010: 56–58) that many forms of collective intentionality, and in particular the forms of collective recognition that are required for the existence of institutional facts, *can* be reductively analysed in individualistic terms. But the main difference between the old and new accounts has to do with the way in which Searle proposes to combine theses (a) and (b) above in providing an explanation of an institutional fact's creation. On the old account, the creation of institutional facts was invariably supposed to be the immediate result of the collective acceptance, within a community, of linguistically expressible "constitutive rules" that specify conditions under which status functions of various sorts are assignable to entities of various sorts. On the new account, the collective acceptance of such constitutive rules remains one but is not the only source from which institutional reality is supposed to spring; the other supposed source is the collective acceptance of speech acts of declaration, whereby entities come to possess status functions just by being represented as having those functions, even in the absence of antecedently available constitutive rules regulating those functions' assignment.

Searle, however, believes that there must be a single fundamental principle underlying the creation of all institutional facts, and so does not rest content with these two seemingly disparate sources. In an attempt to theoretically unify them, he recalls his assumption that, in order to be institutionally effective, constitutive rules have to be linguistically expressible, and reinterprets linguistically expressed constitutive rules that are institutionally effective as *themselves* speech acts of declaration of a special kind, namely as "standing declarations" (Searle 2010: 13, 96–97). He therefore concludes that all of institutional reality derives from the collective acceptance of declarational speech acts (whether "standing" or not "standing" ones) whose effect is the assignment of status functions to entities (and he accordingly calls the acts in question "status function declarations"). Searle clearly regards his proposal to provide a unified account of institutions by tracing their origins to declarational speech acts as the book's most significant contribution: "The main theoretical innovation of this book", he writes, "and one, though not the only, reason for my writing it is that I want to introduce a very strong theoretical claim. All institutional facts, and therefore all status functions, are created by speech acts of a type that in 1975 I baptised as 'Declarations'" (Searle 2010: 11). This claim—and, given the claim's centrality, the book as a whole—appears to me problematic in at least three respects.

The first problem stems from the fact that the theoretical unification the claim strives to achieve is more apparent than real, even assuming that the notion of a

declaration (that is, of a speech act that makes something the case just by representing it as being the case) has no features in Searle's theory of speech acts that prevent its legitimate use in his theory of institutions.[1] The crucial thing to notice in this connection is that, given Searle's own assumption that every institutional fact requires the assignment of a status function to an entity, so-called "standing" status function declarations, unlike status function declarations of the ordinary variety (which I shall henceforth call "grounded" status function declarations), do *not* create institutional facts but only *conditions* for the creation of institutional facts. They do not, in other words, make it the case, just by representing it as being the case, that a particular entity *has* a particular status function. They only make it the case, just by representing it as being the case, that, *if* any entity satisfies such and such conditions, *then* that entity will have such and such a status function.

For example, a community might have collectively accepted a constitutive rule ("standing declaration") specifying under what conditions two individuals belonging to it and married to each other may divorce. But if no married individuals belonging to that community ever seek or obtain the status function of being divorced, no institutional *facts* of divorce will ever exist in that community (and, significantly for Searle, no divorce-dependent rights and obligations will ever be held by any member of that community), even though a well-defined procedure for creating such facts (and such rights and obligations) will be in existence. Similarly, a community might have collectively accepted a constitutive rule ("standing declaration") specifying under what conditions a judge may confer on someone the status function of a person sentenced to death. But as long as no judge actually sentences a person to death by issuing the relevant "grounded" declaration, no institutional *fact* of a person's being sentenced to death will exist in that community (and no member of that community will have such a person's rights and obligations), even though a procedure for creating such facts (and such rights and obligations) will be in place.

Now, if "standing" status function declarations (that is, constitutive rules) do not literally create institutional facts, but, at most, *conditions* for creating institu-

---

1 I will be making that assumption for the sake of argument, but I can easily imagine it being questioned. For example, readers who are aware of the fact that, in earlier works, Searle had spoken of the sorts of declarational acts that he is now taking to be responsible for the existence of extralinguistic institutions as speech acts that "are possible only because of the existence of extralinguistic institutions" (Searle 1998: 150) may worry that there is, at the very least, a threat of circularity lurking behind Searle's current enterprise.

tional facts, the only declaration-invoking instrument left to Searle for accounting, as he intends to do, for the creation of *every* institutional fact is what I have called "grounded" status function declarations. And the problem is that, although some institutional facts can indeed be created by such declarations, many others cannot. A judge can sentence someone to death just by saying, "You are sentenced to death", and an employer can fire an employee just by saying, "You are fired". But I cannot spend money just by saying "I spend money", I cannot go on strike just by saying "I go on strike", and I cannot obey a military order to attack the enemy just by saying to my superior "I obey your order, sir". And since, as Searle would undoubtedly agree, spending money, going on strike, and obeying military orders are institutional facts as much as administering death sentences and firing employees are, it follows that, contrary to his main thesis, not all institutional facts are created by speech acts of declaration.

Ironically, this result was prefigured in a passage of *The Construction of Social Reality* where Searle wrote: "The possibility of creating institutional facts by declaration does not hold for every institutional fact. You cannot, for example, make a touchdown just by saying you are making it" (Searle 1995: 55). That is correct, but it clearly contradicts the thesis Searle is now propounding. And, as I have just argued, the new thesis cannot be relevantly defended by saying that, although one cannot make a touchdown by a "grounded" declaration, there nevertheless must exist somewhere a "standing" declaration (constitutive rule) that lays down conditions under which touchdowns are possible. For, apart from the fact that Searle now accepts that there can be institutional facts not underwritten by any constitutive rules (Searle 2010: 23), merely laying down conditions under which facts of a certain sort are possible does not ensure that any fact of that sort (be it a divorce, a death sentence, or a touchdown) is actual.

The second problem is that Searle's central claim that every institutional fact is brought into existence by the performance of a speech act of declaration sits uncomfortably with his parallel requirement that, in order for institutional facts to exist, they must be collectively *accepted* as existing (Searle 2010: 85–86, 88, 102–103, 106). Searle appears to believe that the two ideas can be seamlessly combined by taking the acceptance requirement on institutional facts to concern the acceptance of relevant speech acts of declaration (where, presumably, accepting a declaration is supposed to amount not to just disinterestedly accepting that a certain kind of utterance has occurred, but rather to accepting *what is said* by that utterance—in other words, accepting the utterance's *propositional content*).

The problem with the proposed combination, however, is that it renders Searle's account of institutional facts incapable of making any essential use of

declarational acts, as Searle has defined them. Recall that a declaration was defined by Searle as a speech act whose speaker (in contrast to the speaker of a mere assertion) makes something the case just by representing it as being the case. It follows, then, that if a speaker's utterance does not make something the case *just* by representing it as being the case, but is rather such that what it represents as being the case cannot become the case *unless* other persons (and indeed, unless a whole community) *accept* that it is in fact the case, then the utterance simply isn't, given the definition, the utterance of a declaration. And since, given the acceptance requirement, the only so-called 'declarations' that Searle would be entitled to appeal to in his account of institutions would be utterances of this last sort, it looks as if no genuine acts of declaration at all could figure in Searle's purported explanation of the creation of institutional reality. Notice that the issue here is far from terminological: it is one thing to claim, for example, that simply by saying, "As of this moment, you have no right to enter this place", a speaker can make it the case that his hearer does not have a certain right, independently of what the hearer's attitudes might be; it is quite another to claim that, by saying, "As of this moment, you have no right to enter this place", a speaker cannot make it the case that his hearer does not have a certain right *unless the hearer accepts* that she doesn't have it.

The problem might be avoidable in a disjunctive account where *some* institutional facts would be taken to be purely acceptance-dependent and some *other* institutional facts would be taken to be purely declaration-dependent. However, Searle's commitment to offer a unitary account of institutional facts prevents him from even considering the possibility of a disjunctive account, and this obliges him to face a dilemma: he must either maintain the acceptance requirement on institutional facts but deny that declarational acts have an essential role to play in their creation, or maintain the view that such acts are essentially involved in the creation of institutional facts, but drop the idea that the facts in question cannot exist unless collectively accepted as existing. As far as I can see, the book contains no clue as to how this dilemma could be avoided.

The third problem is that, in order to preserve his central claim's speech act theoretic commitment, Searle is finally forced to make the scope of his account of institutional facts significantly narrower than initially advertised, and to thereby abandon what he was taking the guiding methodological principle of his project to be. As already noted, Searle begins by presenting the book's "main theoretical innovation" as the "very strong theoretical claim" that "all institutional facts" are created by declaration (Searle 2010: 11). Soon afterwards, however, Searle switches to a heavily qualified version of that claim: he now claims that all of institutional reality "with the important exception of language itself" (Searle

2010: 12) is created by declaration—"language itself", he later points outs, is an institutional phenomenon that "is not created by declaration" (Searle 2010: 110), and the hypothesis that it is so created should be rejected on account of necessitating "an infinite regress" (Searle 2010: 111). Taken together with Searle's unqualified affirmation that, nevertheless, "language is the fundamental social institution" (Searle 2010: 113), the heavily qualified version of the "very strong claim" entails that all institutional facts *except the fundamental ones* are created by declaration—which, I presume, is what we must take Searle's final position on the matter to be.

One might at this point ask how, if that is indeed Searle's final position, he could end the book by claiming, in what appears to be a confident reiteration of his opening claim, that "all of human institutional reality" is "created in its initial existence and maintained in its continued existence" by the "single logico-linguistic operation" of "a Status Function Declaration" (Searle 2010: 201). The answer to this question is that, halfway through the book, Searle has issued the following warning: "From now on, when I say 'institutional fact' I mean 'non-linguistic institutional fact'" (Searle 2010: 93). Searle's closing claim, then, is only typographically a reiteration of his opening one. In reality, it is not a claim about "all of human institutional reality" at all, but only a claim about that part of human institutional reality—the non-linguistic one—that, according to Searle, is *not* the fundamental part.

Now, the two problems previously raised suggest that, even in this considerably weakened form, Searle's position would be unsatisfactory. But independently of that, the fact that the position Searle is finally led to embrace is so weakened should itself have given him pause, in view of his strongly voiced opinion, presented as the guiding methodological principle of his enterprise, that it is "implausible to suppose that we would use a series of logically independent mechanisms for creating institutional facts" (Searle 2010: 7). If that is implausible, then embracing the weakened position is itself implausible, since embracing it amounts to accepting that, apart from the declarational mechanism allegedly responsible for the creation of non-linguistic institutions, there must be one or more logically independent (and so far unspecified) mechanisms responsible for the creation of linguistic ones. It would seem, then, that anyone who takes the extent of institutional reality to be the same as Searle takes it to be—and, in particular, anyone who, like Searle, takes linguistic facts to *be* institutional facts—, and who, at the same time, is committed to finding the "single mechanism" (Searle 2010: 7) that underlies *every* institutional fact's creation—which is what Searle was searching in this book—, must begin a new search.

Let us hope that Searle's next book will undertake such a search. In the meantime, the present one may be recommended to newcomers to his philosophy as a lively, if occasionally repetitive, introductory overview of many of his current research themes and of some of his past research achievements.

Chapter 21
A paradox of cooperation in the theory of implicatures

1 Introduction

That the utterance of a sentence may be intended to convey, and may succeed in conveying, propositions that are in no sense part of what would be regarded as the sentence's conventional meaning is a thesis that no one would be inclined to seriously dispute. But the question as to how uttered sentences manage to convey such propositions is a question to which no one before Grice (1975, 1989) has attempted to give a comprehensive answer. Since there is a significant amount of linguistic and philosophical work where the overall success of the Gricean attempt is taken for granted, it is important to know whether it really succeeds. My purpose in this essay is to present a single—but, I believe, crucial—type of counterexample suggesting that it doesn't.

Grice calls "conversational implicatures" those communicated propositions that are not part of what would be regarded as the conventional meanings of the sentences whose utterance communicates them; and he claims that, even independently of the question as to how conversational implicatures manage to be communicated, it would be reasonable for us to assume that, in the absence of indications to the contrary, participants in a talk exchange follow, and expect each other to follow, the general principle (called "Cooperative Principle") specified in (I), under which the various sub-principles (called "Conversational Maxims") specified in (II) may be regarded as falling:

(I) "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged." (Grice 1989: 26)

(II) "Make your contribution as informative as required"; "Do not make your contribution more informative than is required"; "Do not say what you believe to be false"; "Do not say that for which you lack adequate evidence"; "Be relevant"; "Be perspicuous". (Grice 1989: 26–27)

Apart from being independently plausible, Grice then contends, the assumption that conversationalists respect, and expect each other to respect, the Cooperative Principle provides the essential ingredient of an explanation of the fact that the utterance of a sentence whose conventional meaning goes no further than the proposition *p* may nevertheless succeed in conversationally implicating the further proposition *q*: that further proposition will be conversationally implicated, Grice asserts, whenever the hypothesis that it is intended to be conveyed by the speaker is the hearer's best grounds for maintaining the presumption that the speaker *has not abandoned the Cooperative Principe* (a presumption that is maintained, according to Grice, both in those cases in which the hearer thinks that the speaker, in saying what he does, observes each one of the Maxims associated with the Principle and in those cases in which the hearer notices that the speaker, in saying what he does, violates in a *blatant* way—that is, in a way that is intended to be recognised and cannot fail to be recognised—one or another of the Maxims associated with the Principle). Specifically, a speaker who has said that *p* will be taken, according to Grice, to have conversationally implicated the further proposition *q*, provided that his hearer will have reasoned, and will have been intended to reason, as follows:

(III)   "He has said that *p*; there is no reason to suppose that he is not observing the maxims, or at least the cooperative principle; he could not be doing this unless he thought that *q*; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that *q* IS required; he has done nothing to stop me thinking that *q*; he intends me to think, or is at least willing to allow me to think, that *q*; and so he has implicated that *q*." (Grice 1989: 31)

Thus—to take two well-known kinds of example—the fact that I can convey the proposition that not all of my students are clever by saying, "Some of my students are clever", is to be explained, according to Grice, by supposing that the hypothesis that I think that not all of my students are clever is my hearer's best grounds for thinking that, in saying what I said, I have *not abandoned the Cooperative Principle*, and have, in particular, observed the maxim, "Make your contribution as informative as required." And the fact that I can convey the proposition that a certain person is particularly unkind by saying of him, "He is a pig", is to be explained, according to Grice, by supposing that the hypothesis that I think of that person as particularly unkind is my hearer's best grounds for thinking that, in saying what I said, I have *not abandoned the Cooperative Principle*, even though

(given that it is obvious that a person cannot literally be a pig) I have violated in a *blatant* way the maxim, "Do not say what you believe to be false."

In what follows I will not attempt, as some commentators have done (see, for example, Kasher 1976, Sampson 1982, Sperber and Wilson 1986, Taylor and Cameron 1987), to answer the question whether conversationalists are or are not in fact trying to be cooperative. My argument will simply be that they cannot both be cooperative in the sense that Grice supposes them to be *and* follow the kind of reasoning that Grice thinks they follow in calculating conversational implicata. This, I shall argue, is a consequence of the fact that there is a particular variety of conversational implicatures—which I shall here call "implicatures of non-cooperation"—such that the supposition that conversationalists derive implicatures of that kind by instantiating the inference schema that Grice specifies in (III) is inconsistent with the grounds that Grice offers for supposing that they adhere to the general principle specified in (I).

2 Implicatures of non-cooperation

What is Grice's reason for supposing that participants in a conversation behave, and expect each other to behave, in accordance with the Cooperative Principle specified in (I)? Grice's schematic answer to this "fundamental question" (Grice 1989: 28) comprises two claims. The first claim is that, since it is distinctive of intentional human behaviour to be *rational* behaviour, and since conversation is a characteristic form of intentional human behaviour, conversational behaviour can be presumed to be "a special case or variety" of "rational behaviour" (Grice 1989: 28). And the second claim is that conversational behaviour would not be rational behaviour *unless* those engaging in it behaved, and expected each other to behave, in accordance with the Cooperative Principle: "observance of the Cooperative Prince", Grice tells us, is the central requirement on anyone aspiring to be a "rational" conversational participant, in that "anyone who cares about the goals that are central to conversation/communication (such as giving and receiving information, influencing and being influenced by others) must be expected to have an interest, given suitable circumstances, in participation in talk exchanges that will be profitable *only on the assumption* that they are conducted in general accordance with the Cooperative Principle" (Grice 1989: 29–30; italics added).

Given that the rationality of speech transactions essentially involves, for Grice, the observance, on the part of each conversational participant, of the principle specified in (I), it is easy to extract from the above schematic answer the

following minimal conditions on conversational rationality: First, that, *ceteris paribus,* it would not be rational for a conversational participant to act (or to keep acting) in accordance with principle (I) unless he believed that his conversational partners are acting *themselves* (or are trying to act *themselves)* in accordance with principle (I). And second, that, *ceteris paribus,* it would not be rational for a conversational participant to believe that his conversational partners *are* acting (or *are* trying to act) in accordance with principle (I) unless he believed that *they* believe *him* to be acting (or to be trying to act) in accordance with principle (I). The mutual dependence, highlighted in these conditions, of the attitudes that, according to Grice, sustain conversational rationality is prominent in, among others, the following passage: "Our talk exchanges do not normally consist of a succession of unconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts: and each participant recognises in them, to some extent, a common purpose or set of purposes, or at least a mutually accepted direction" (Grice 1989: 26). In that passage, the claim that conversational activities are rational is spelled out as the claim that their protagonists have mutually recognised common goals, and make collaborative—or, as Grice also says, "dovetailed, mutually dependent" (Grice 1989: 29)—efforts towards achieving those common goals. Given that the efforts in question take, as Grice supposes, the form of attempts to observe the general principle specified in (I), and given that these attempts could hardly be taken to be 'dovetailed' unless the observance of principle (I) by any one participant was dependent on his assumption that the other participants observe (or try to observe) it too, and that they would not be observing (or be trying to observe) it unless they assumed that *he* is observing (or is trying to observe) it too, it follows that any claim to the effect that the principle's observance is rational entails, for Grice, the claim that these assumptions are in fact being made by those who are observing it.

Now, anyone who wishes to claim that observance of the Cooperative Principle is a practice that is maintained because it is rational in the above sense, must be prepared to admit that the Gricean explanation of conversational implicatures by reference to the inference schema in (III) would be revealed to be unsuccessful if it turned out to be the case that what a speaker can *conversationally implicate* is to the effect that his interlocutor *is not observing the Cooperative Principle*. For, according to any explanation that essentially involves the inference schema in (III), no hearer is entitled to regard *any* proposition as conversationally implicated *unless* he regards the hypothesis that it is accepted by the speaker as his best grounds for thinking that the speaker *observes the Cooperative Principle*. And since, if a hearer is rational in the sense just explicated, he can hardly regard the

hypothesis that the speaker takes him *not* to be observing the Cooperative Principle as his best grounds for thinking that the speaker himself *is* observing the Cooperative Principle, it would follow that the implicature in question could be derived by a user of the Gricean derivational mechanism *only if* that user was irrational by Gricean standards.

I submit that implicatures of the sort just described are perfectly possible and indeed not uncommon. Suppose that A and B have a conversation on a certain topic (whose precise nature need not concern us); suppose further that, at a certain point in the course of that conversation, A starts posing to B a series of questions that are quite relevant to one aspect of the conversation's topic; and finally suppose that B, either because he wants to avoid taking a stand on that particular aspect of the topic or because he is distracted by some external events, remains silent throughout A's attempts to elicit from him answers to his questions. Exasperated by B's unresponsiveness, A then emits the following memorable utterance:

(1)   I wonder why I am talking to you.

One thing that is clear about (1) is that its utterance in the context just described may be intended to convey, and may fully *succeed* in conveying, the proposition that B, the hearer, *is not making his conversational contribution such as is required by the accepted purpose or direction of the talk exchange he is engaged in*—in other words, the proposition that B is *not* being conversationally cooperative (of course, there are many further propositions that the utterance of (1) can successfully convey in this context, but it can certainly convey this one, and that is the important point for the present discussion). And one thing that is clear about that conveyed proposition is that it is in no sense part of what (1) conventionally *means*—in other words, that, no matter what theory of 'conventional meaning' you choose, it will never tell you that "I wonder why I am talking to you" *means the same thing as* "You are not making your conversational contribution such as is required by the accepted purpose or direction of the talk exchange you are engaged in" (after all, the state of affairs in which every speaker acts in full conformity with the agreed canons of conversational propriety is perfectly consistent with the state of affairs in which some among those speakers keep wondering why such things as conversations occur at all). If this is so, the conveyed proposition in question must surely be counted as a *conversational implicature* of the utterance of (1) in the context described, and the question must arise as to whether the Gricean schema in (III) does or does not enable the addressee of (1) to infer the implicature's presence.

To see that the inference is impossible unless the addressee of (1) explicitly contradicts the Gricean conception of conversational rationality, it is sufficient to inspect the relevant substitution instance of schema (III), which is constructed below by letting '*s* observes the cooperative principle' stand as an abbreviation of '*s* makes his/her conversational contribution such as is required by the accepted purpose or direction of the talk exchange in which *s* is engaged':

(X)   He has said that he wonders why he is talking to me; there is no reason to suppose that he is not observing the cooperative principle; he could not be doing this unless he thought that I am not observing the cooperative principle; he knows (and knows that I know that he knows) that I can see that the supposition that he thinks that I am not observing the cooperative principle IS required; he has done nothing to stop me thinking that I am not observing the cooperative principle; he intends me to think that I am not observing the cooperative principle; and so he has implicated that I am not observing the cooperative principle.

Although almost every step of this inference has its fair share of paradox, the most important one to notice for present purposes is the third, which deserves to be formulated slightly more fully:

(x)   He could not be observing the cooperative principle unless he thought that I am not observing the cooperative principle.

The problem with (x) is that if, as Grice's derivational mechanism requires, it does occur as part of B's reasoning about A's utterance of (1), then it implies that, contrary to what Grice would expect, B has abandoned all *rational* grounds for thinking that A upholds the Cooperative Principle. For, Grice would be the first to insist that it would *not* be rational on B's part to think that A upholds the Cooperative Principle unless he thought that A thinks that B *himself* is upholding the Cooperative Principle (this, to repeat, is what would make the observance of the principle a *cooperative* enterprise in the first place). However, B's train of thoughts is certainly incapable of respecting that rationality requirement *if* it is to conform to the inference schema that Grice regards as responsible for the generation of conversational implicatures. For, on that schema, every implicature that a hearer derives must be such that the supposition that it is part of the speaker's thoughts is the hearer's indispensable premise for inferring that the speaker *is* upholding the Cooperative Principle. And since one of the implicatures that A happens to be able to convey by uttering (1) is that B is *not* upholding the Cooperative Principle,

the derivation of that implicature by means of the schema certainly requires B to think, no matter how irrationally, that A would uphold the Cooperative Principle *only if* he was taking B *not* to uphold it. So, either Grice was wrong in supposing that speakers and hearers calculate conversational implicatures in the way he suggests, or he was wrong in supposing that speakers and hearers observe the Cooperative Principle for the reasons he offers.

3 Conclusion

There are, so far as I can see, only two options available to someone who would be in a position to recognise the seriousness of this problem, but who would not wish to abandon in its entirety the complex of Gricean ideas from which it springs. On the one hand, one might retain the account of implicature calculation proposed by Grice, but in that case one would have to claim, *contra* Grice, that the principle on whose presumed observance the intelligibility of the account depends is only in name, and not in fact, a Cooperative Principle, that no one really has any *rational* grounds for respecting rather than for not respecting it. On the other hand, one might retain the Gricean thesis that the Cooperative Principle is a principle whose observance, whenever it occurs, constitutes an eminently rational form of behaviour, but in that case one would have to claim, *contra* Grice, that the presumption of its observance plays no role whatsoever in any inferences responsible for the calculation of conversational implicatures. I suspect that no linguist or philosopher among those who have confidently proposed Gricean explanations of their data would find either of these options satisfactory. But I am afraid that if the options are unsatisfactory, then so are the proposed explanations.

Chapter 22
An inferential impasse in the theory of implicatures

1 Introduction

The goal of the Gricean theory of conversational implicatures (Grice 1975, 1989), and of several of its contemporary descendants, is to explain how utterances of sentences can be intended to convey, and can be understood as conveying, information lying beyond what they are semantically equipped to convey. The purpose of this essay is to show that the Gricean explanation of two prominent types of putative conversational implicatures faces a so far unnoticed problem when confronted with utterances that simultaneously carry implicatures of *both* of these prominent types. The problem, in a nutshell, is that, since the Gricean theory requires implicatures of these two types to be calculated under mutually incompatible inferential regimes, it cannot without inconsistency derive implicatures of either type when a single utterance carries both of them. After explaining how this problem—which I will call "the problem of composite implicatures"— arises, the essay briefly indicates why it would fail to arise if certain distinctively anti-Gricean, but independently supported, assumptions about utterance interpretation were adopted.

2 The problem of composite implicatures

On the Gricean theory, conversational implicatures are conclusions of inferences that hearers draw, and are intended by speakers to draw, in order to maintain the presumption that speakers, in making their utterances, respect the so-called Cooperative Principle, a principle that is taken to embody the view that "talking [is] as special case or variety of...rational behavior" (Grice 1989: 28) and which Grice formulates as the prescription, "Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged" (Grice 1989: 26). Furthermore, the inferences in question are supposed to belong to two fundamentally different types. In inferences of the first type, hereafter called *Observance-Induced* infer-

ences, the hearer is supposed to derive the implicature by assuming (and assuming that she is expected to assume), on the one hand, that the speaker respects the Cooperative Principle and, on the other hand, that he does not contravene any of the so-called Conversational Maxims associated with that Principle (the Maxims of Quality, Quantity, Relation, and Manner; Grice 1989: 26–27). In inferences of the second type, hereafter called *Violation-Induced* inferences, the hearer is supposed to derive the implicature by assuming (and assuming that she is expected to assume), on the one hand, that the speaker respects the Cooperative Principle, and, on the other hand, that he *blatantly* contravenes—that is, contravenes in a way that is meant to be recognized and cannot fail to be recognised—at least one of the so-called Conversational Maxims associated with that Principle. It will be noticed that since inferences of these two types incorporate obviously inconsistent assumptions—it is impossible both to contravene at least one of the Maxims and to not contravene any of them—, a minimally rational hearer could not be supposed to be jointly employing inferences of the two types in calculating any given conversational implicature.

Classic examples of utterance interpretations that, on Gricean theories, are conversational implicatures communicated by virtue of Violation-Induced inferences are the interpretations that involve recognising what are traditionally called "figures of speech"—in particular, metaphorical or metonymic uses of expressions—, and these will be hereafter called M-implicatures. The fact, for example, that an utterance like (1) will, in many contexts, be taken to convey a proposition such as (Imp-1),

(1)   John's suggestion was shot down.
(Imp-1) John's suggestion was rejected.

or the fact that an utterance like (2) will, in many contexts, be taken to convey a proposition such as (Imp-2),

(2)   Mary's guest is a big name.
(Imp-2) Mary's guest is a famous person.

are supposed to be explicable by reference to inferences that essentially involve the assumption that the utterers of (1) and (2) blatantly violate, in speaking as they do, the Maxim of Quality, which requires of conversational participants to aim at truth ("try to make your contribution one that is true", Grice 1989: 27), and to therefore refrain from saying what they believe to be false ("do not say what you believe to be false", Grice 1989: 27). Given that it is obviously false, and hence

unlikely to be believed by anyone, that a suggestion can be the object of a shooting, or that a guest can be a name, hearers of (1) and (2) are supposed to be taking speakers of these utterances to be blatantly violating the Maxim of Quality; but since these same hearers are also supposed to be taking the speakers to be respecting the Cooperative Principe, they are held to have the task of resolving the clash between their interlocutors' blatant violation of the Maxim and their interlocutor's presumed adherence to the Principle; and, according to the Gricean account, hearers resolve that clash by inferring that what their interlocutors intend to convey, and to be understood as intending to convey, by their utterances is not what those utterances semantically express but something different from what they semantically express—for example, in the case of (1) that the suggestion to which the speaker is referring was rejected, and in the case of (2) that the guest to which the speaker is referring is a famous person.

Classic examples of utterance interpretations that, on Gricean theories, are communicated by virtue of Observance-Induced inferences are the interpretations that are nowadays commonly referred to as "scalar implicatures", and which will hereafter be called S-implicatures. For example, the fact that an utterance like (3) will, in many contexts, be taken to convey a proposition such as (Imp-3),

(3)   Some of John's suggestions were rejected.
(Imp-3)   Not all of John's suggestions were rejected.

or the fact that an utterance like (4) will, in many contexts, be taken to convey a proposition such as (Imp-4),

(4)   Some of Mary's guests are famous persons.
(Imp-4)   Not all of Mary's guests are famous persons.

are supposed to be explicable by reference to inferences that essentially involve the assumption that the speakers of (3) and (4) respect not only the Cooperative Principle but also each one of the Conversational Maxims associated with that Principle, and in particular the Maxims of Quality and Quantity. The Gricean hearer's reasoning in these cases is supposed to proceed, and to be intended by the speaker to proceed, as follows. By virtue of assuming that the speaker of (3) or (4) respects the Maxim of Quality, the hearer infers that the speaker does believe in the truth that what is semantically expressed by the sentence he utters. By assuming, furthermore, that the speaker of (3) or (4) also respects the Maxim of Quantity, she infers that what is semantically expressed by the sentence he

utters is the maximum amount of relevant information that he considers himself entitled to convey while remaining truthful. From this, the hearer infers that the speaker does not believe in the truth of what would be expressed by a certain semantically stronger sentence that he might have relevantly uttered but did not in fact utter. And from this she concludes that he invites, or at least allows, her to think that what would be expressed by that semantically stronger sentence is not true. Thus, on the assumption that the speaker of (3) is both truthful in what he says (as required by Quality) and maximally informative in what he says (as required by Quantity), his choosing to say not that all of John's suggestions were rejected, but only that some of them were rejected, leads the hearer to suppose that he does not believe that all of them were rejected, and is therefore inviting, or at least allowing, her to think that not all of them were rejected. And similarly, on the assumption that the speaker of (4) is truthful in what he says (as required by Quality) and maximally informative in what he says (as required by Quantity), his choosing to say not that all of Mary's guests are famous persons, but only that some of them are, leads the hearer to suppose that he does not believe that all of them are, and is therefore inviting, or at least allowing, her to think that not all of them are. It is worth emphasising, and it is emphasised by Grice himself (1989: 27, 371), that the assumption that a speaker, in saying what he does, is being maximally informative, and so respects the Maxim of Quantity, only makes sense if it is antecedently assumed that, in saying what he does, he is being truthful, and so respects the Maxim of Quality—the reason, as Grice (1989: 371) notes, is that "false information is not an inferior kind of information; it just is not information."

Now, a theory that purports to derive, in the Gricean way just sketched, M-implicatures via Violation-Induced inferences and S-implicatures via Observance-Induced inferences would be shown to be inadequate if there could be utterances each of which would simultaneously carry an M-implicature and an S-implicature. For, if utterances carrying such composite implicatures existed, deriving the implicatures in the Gricean way would require the joint employment of a Violation-Induced inference and an Observance-Induced inference; and these two types of inference could not possibly be co-opted by a rational hearer—as Gricean hearers are supposed to be—, since co-opting them would require accepting as true the obvious contradiction that a speaker simultaneously violates at least one of the Conversational Maxims and does not violate any of them.

Unfortunately for the Gricean program, it is an easily ascertainable fact that composite implicatures of the sort just described are perfectly possible, and indeed quite common. For example, responding to an utterance like (5),

(5)   Have John's suggestions been rejected?

it is perfectly possible for a speaker to utter the sentence in (6),

(6)   Some of John's suggestions have been shot down.

and to thereby convey the implicature in (Imp-6):

(Imp-6)   Not all of John's suggestions have been rejected.

And responding to an utterance like (7),

(7)   Is it true that Mary's guests are famous people?

it is perfectly possible for a speaker to utter the sentence in (8),

(8)   Some of Mary's guests are big names.

and to thereby convey the implicature in (Imp-8):

(Imp-8)   Not all of Mary's guests are famous people.

But the derivation of these composite implicatures is not possible on Gricean assumptions. In order to derive any S-implicature, and so the implicature involving the transition from "some of John's suggestions" to "not all of John's suggestions" in the case of (6), or the one involving the transition from "some of Mary's guests" to "not all of Mary's guests" in the case of (8), the Gricean hearer must employ an Observance-Induced inference; and that inference, as we saw, must begin with the hearer's assumption that the speakers of these utterances do respect the Maxim of Quality, and therefore *do believe that what is semantically expressed by the sentences they utter is true*. In order, on the other hand, to derive any M-implicature, and so the implicature involving the transition from "were shot down" to "were rejected" in the case of (6), or the one involving the transition from "are big names" to "are famous people" in the case of (8), the Gricean hearer must employ a Violation-Induced inference; and that inference, as we saw, must begin with the hearer's assumption that the speakers of these utterances blatantly violate the Maxim of Quality, given that they obviously *do not believe that what is semantically expressed by the sentences they utter is true*. However, it is incoherent to suppose that anyone can simultaneously violate and not

violate the Maxim of Quality, and equally incoherent to suppose that anyone can simultaneously believe and not believe that what is semantically expressed by a certain uttered sentence is true. Therefore, a rational hearer using the resources offered her by the Gricean theory can derive neither the composite implicature communicated by (6) nor the composite implicature communicated by (8)—nor, of course, any of the many composite implicatures of a similar kind that utterances of natural language sentences can easily convey.[1]

One might try to ignore the problem just outlined by declaring that the M-implicatures and S-implicatures jointly conveyable by utterances of (6) or of (8) are, despite all appearances, creatures of an entirely different kind from the M-implicatures or S-implicatures conveyable by utterances of (1)–(2) or of (3)–(4), and that implicatures of that mysterious new kind lie outside the intended scope of the Gricean theory of implicatures. This is the sort of response that would protect the Gricean theory only by making it unfalsifiable, and so would not save it at all. In addition, it would be a response with extremely implausible implications for the particular cases under consideration: surely, if any mechanism explains the transition from "some" to "not all" in the examples that appear unproblematic for the Gricean account (such as (3) and (4)), the *same* mechanism should be able to explain the identical transition from "some" to "not all" in the examples that are demonstrably problematic for that account (such (6) and (8)); and similarly, if any mechanism explains the transitions from "was shot down" to "was

---

1 It should be noticed, in case it isn't obvious, that the idea of a two-phased derivation of composite implicatures, in which the output of an initial Violation-Induced inference would be the input to a subsequent Observance-Induced inference, would make no Gricean sense (assuming that it would make sense at all). On such a proposal, the hearer of "Some of John's suggestions have been shot down" would first conclude, via a Violation-Induced inference, that the speaker believes that some of John's suggestions have been rejected, and then, by treating *her own* conclusion *as if* it was a further utterance by the speaker, and by applying to that *imaginary* speaker's utterance an Observance-Induced inference, would derive the conclusion that the actual speaker does not believe that all of John's suggestions have been rejected; and similarly, the hearer of "Some of Mary's guests are big names" would first conclude, via a Violation-Induced inference, that the speaker believes that some of Mary's guests are famous persons, and then, by treating *her own* conclusion *as if* it was a further utterance by the speaker, and by applying to that *imaginary* speaker's utterance an Observance-Induced inference, would derive the conclusion that the actual speaker does not believe that all of Mary's guests are famous persons. The trouble with this proposal is, of course, that Gricean inferences are supposed to be triggered by actual speakers' utterances rather than by imaginary speakers' utterances; and that, in any case, it is only regarding the former, and not the latter, type of utterance that the question as to what it implicates, and whether or not it respects any conversational principles, can significantly be answered, and so can significantly be asked**.**

rejected", or from "is a big name" to "is a famous person", in the examples that appear unproblematic for the Gricean account (such as (1) and (2)), the *same* mechanism should be able to explain the analogous transitions in the examples that are demonstrably problematic for that account (such as (6) and (8)). Instead, therefore, of trying to protect the Gricean doctrine by making it unfalsifiable, it would be advisable to admit that the composite implicatures conveyable by utterances like (6) and (8) do falsify it, and then consider which theoretical decisions would be consistent with recognition of this fact.

One kind of solution would be to abandon the Gricean explanation of M-implicatures in terms of Violation-Induced inferences and retain *only* the Gricean explanation of S-implicatures in terms of Observance-Induced inferences (assuming that the latter type of explanation would not be objectionable on independent grounds). On one option within that kind of solution, the interpretation of metaphorical or metonymic uses of expressions that an utterance may contain would be part of a contextually adjusted extension of its *semantic* interpretation—in other words, what are here called M-implicatures would not be instances of conversational implicature in Grice's sense of that term, and would not be 'explained' by Violation-Induced inferences of the Gricean kind—, whereas S-implicatures, which would *be* instances of conversational implicature in the Gricean sense, would be pragmatically derived when needed, and with no risk of inconsistency, through Observance-Induced inferences of the Gricean kind.

A second kind of solution would be to abandon the Gricean explanation of S-implicatures in terms of Observance-Induced inferences and retain *only* the Gricean explanation of M-implicatures in terms of Violation-Induced inferences (assuming that the latter type of explanation would not be objectionable on independent grounds). On one option within that kind of solution, the derivation of an utterance's so-called "scalar implicatures" would be part of a contextually adjusted extension of its *semantic* interpretation—in other words, what are here called S-implicatures would not be instances of conversational implicature in Grice's sense of that term, and would not be 'explained' by Observance-Induced inferences of the Gricean kind—, whereas M-implicatures, which would *be* instances of conversational implicature in the Gricean sense, would be pragmatically derived when needed, and with no risk of inconsistency, through Violation-Induced inferences of the Gricean kind.

Finally, a third kind of solution would be to abandon both the Gricean explanation of M-implicatures in terms of Violation-Induced inferences *and* the Gricean explanation of S-implicatures in terms of Observance-Induced inferences. On one option within that kind of solution, the interpretation of both the metaphorical or metonymic uses of expressions that an utterance may contain *and* the

so-called "scalar implicatures" that an utterance may convey would be parts of a contextually adjusted extension of its *semantic* interpretation, and only aspects of communicated content *beyond* those here referred to as M-implicatures or S-implicatures could possibly be pragmatically analysed in Gricean terms.

It is not my purpose here to argue in favour of one of these types of solution and against the others, but it is clear that each type, and especially the third one, would require acknowledging that the Gricean theory is significantly less successful than its supporters take it to be: Gricean explanations of S-implicatures in terms of Observance-Induced inferences, as well as Gricean explanations of M-implicatures in terms of Violation-Induced inferences, have long been cited, and keep being widely cited, as prime examples of the explanatory fertility of the Gricean approach to the analysis of linguistic interpretation; therefore, if, as the problem of composite implicatures discussed here suggests, at least one, and possibly both, of these types of Gricean explanation cannot be sustained, the range of theoretically interesting cases that the Gricean theory can adequately cover would appear to be drastically reduced.

The significance of this outcome would of course be enhanced if there were theoretical frameworks, alternative to the Gricean one, and developed on the basis of evidence independent of the evidence provided by composite implicatures, for which the existence of composite implicatures would be in principle unproblematic. And it seems to me that there are in fact several such frameworks, two of which I would now like to briefly mention, concentrating on what would make their treatment of composite implicatures anti-Gricean rather than on the important differences that exist between them.

As a first example, consider the treatment of what are here called S-implicatures and M-implicatures within Relevance Theory (Sperber and Wilson 1995; Carston 2002, 2010; Wilson and Sperber 2012). On that theory, the interpretative phenomena associated with these labels are not implicatures properly so called but rather *explicatures*—that is, representations of the explicit content of an utterance in a context—that are derived from an utterance's possibly underspecified logical form in accordance with the Principle of Relevance; and neither explicatures nor implicatures properly so called—that is, representations of the non-explicit content of an utterance in a context, also derived in accordance with the Principle of Relevance—owe their existence to speaker intentions aiming to provoke hearer inferences that invoke the Cooperative Principle and the observance or violation of Gricean Conversational Maxims. Now, an explicature, according to Relevance Theory, can involve (among other things) either the *narrowing* or the *broadening* of the conceptual content of an element present in an

utterance's logical form; and what are here called S-implicatures and M-implicatures, when occurring independently of each other, could be analysed as components of explicatures resulting, respectively, from a conceptual narrowing process applied to an element present in an utterance's logical form and from a conceptual broadening process applied to an element present in an utterance's logical form. Thus, the S-implicatures involving the transition from "some of John's suggestions" to "not all of John's suggestions" in the interpretation of (3), or the transition from "some of Mary's guests" to "not all of Mary's guests" in the interpretation of (4), could be represented as components of explicatures resulting from the application of a conceptual *narrowing* process to appropriate elements in those utterances' logical forms; whereas the M-implicatures involving the transition from "was shot down" to "was rejected" in the interpretation of (1), or the transition from "is a big name' to "is a famous person" in the interpretation of (2), could be represented as explicatures resulting from the application of a conceptual *broadening* process to appropriate elements in those utterances' logical forms. Notice, however, that there is nothing in Relevance Theory that prevents conceptual narrowing and conceptual broadening processes from operating *in parallel*, as long as they operate on *distinct* elements of an utterance's logical form. Consequently, Relevance Theory would have no difficulty in accounting for the fact, which a Gricean theory cannot coherently explain, that the same utterance can simultaneously convey what is here called an S-implicature and an M-implicature, as this happens in (6), in (8), and in many other utterances of a similar kind. For, these composite implicatures would simply be, in relevance-theoretical terms, explicatures resulting from the simultaneous application of a conceptual narrowing process to *one* element of an utterance's logical form and of a conceptual broadening process to a *distinct* element of the same utterance's logical form. And, of course, simultaneously applying a conceptual narrowing process to one element of an utterance's logical form and a conceptual broadening process to a different element of the same utterance's logical form is not the same thing as incoherently supposing that the utterance's speaker simultaneously violates and does not violate a Conversational Maxim, or simultaneously believes and does not believe that what is semantically expressed by the sentence that he or she utters is true.

As a second example, consider theories that have proposed, on the basis of various types of evidence that they consider to be impossible to explain under Gricean assumptions, that the interpretative phenomena here referred to as S-implicatures or M-implicatures do not owe their existence to reflexively intended pragmatic inferences invoking the Cooperative Principle and the observance or violation of Gricean Conversational Maxims, but are rather due to the presence,

in the logical form of the utterances concerned, of dedicated silent operators that determine those utterances' *semantic* interpretation relative to various bodies of contextually available information. In the case of scalar phenomena, an approach of this sort is the so-called "grammatical view of scalar implicatures" developed by Chierchia and others, according to which what have here been referred to as the S-implicatures of utterances like (3) and (4) would be analysed as components of the semantic interpretation of those utterances, resulting from the interaction between a silent exhaustification operator present in the utterances' logical forms and a contextually salient set of alternatives to the proposition embedded under that operator (see Chierchia 2004, 2013, 2017; Chierchia, Fox and Spector 2012). In the case of phenomena like metaphor and metonymy, an approach of this sort is the so-called "demonstrative" account of figurative language initiated by Stern, according to which what have here been referred to as the M-implicatures of utterances like (1) and (2) would be analysed as components of the semantic interpretation of those utterances, resulting from the interaction between a silent "Mthat" operator—in some respects analogous to the Kaplanian "Dthat" operator—present in those utterances' logical forms and a contextually salient set of properties related to the property denoted by the predicate embedded under the "Mthat" operator (see Stern 1985, 2000, 2006, 2011). Theories like Chierchia's and Stern's have typically addressed, in their non-Gricean ways, 'pure' S-implicatures and 'pure' M-implicatures, respectively, without considering cases of utterances, such as (6) and (8) above, that simultaneously convey implicatures of both types. However, there is no good reason to suppose that they could not be combined in order to provide an account of such composite implicatures, and to thereby achieve what Gricean theories cannot achieve. The key to their combinability is the fact that the operators they respectively appeal to have different scopal properties. A Stern-type operator applies, fundamentally, to a predicate-value and delivers a different predicate-value selected from a contextually salient set of predicate-values, whereas a Chierchia-type operator applies to a proposition and delivers a different proposition constructed out of the first and of a contextually salient set of propositional alternatives to the first. In the representation of sentences conveying composite implicatures, then, the Stern-type operator would have narrow scope whereas the Chierchia-type operator would have wide scope—for example, in the representation of (6) and (8), the scopally relevant configurations would be as in (6′) and (8′), respectively, where *Z* stands for the Chierchia-type operator and *Y* for the Stern-type operator:

(6′)   [*Z* [Some of John's suggestions [*Y* [were shot down]]]]
(8′)   [*Z* [Some of Mary's guests [*Y* [are big names]]]]

And the interpretation would proceed, in the standard way, by first computing the effect of the narrow-scope Stern-type operator and by then computing the effect of the wide-scope Chierchia-type operator. The overall interpretation would thus encompass both the M-implicature (as a result of the first computation) and the S-implicature (as an entailment of the result of the second computation). It is clear, however, that representing a composite implicature as the compositional outcome of semantic processes triggered by scopally distinct operators in the logical form of the utterance carrying it is not the same thing as 'explaining' the implicature by supposing that hearers of the utterance carrying it incoherently assume (and are intended by the speaker to incoherently assume) that the speaker both contravenes and does not contravene a Conversational Maxim, or both believes and does not believe that what is semantically expressed by the sentence he or she utters is true.

3 Conclusion

It would be an interesting further task, which I will not undertake here, to examine which one of the two types of non-Gricean approach very briefly sketched in the preceding two paragraphs would offer, when articulated in detail, the best account of composite implicatures. My concern has been simply to argue that, although each of these approaches has been designed to overcome defects of the Gricean approach that are independent of the specific problem of composite implicatures, neither of them is conceptually precluded, as the Gricean approach is, from addressing that problem as well. If that is correct, what have here been described as composite implicatures should be added to the growing list of phenomena that suggest that what may appear as areas of interpretation paradigmatically amenable to a Gricean treatment turn out on closer inspection to be areas that can be adequately approached only if fundamental tenets of the Gricean program are jettisoned.

Part III **Knowledge matters**

Chapter 23
How to forget that "know" is factive

1 An inauspicious undertaking

Knowledge, it is generally supposed, entails truth, and few of the questions that have preoccupied analytic epistemologists over the years would make good sense if it didn't. For this reason, serious attempts to dispute that knowledge entails truth deserve scrutiny, and my purpose in what follows is to scrutinise one such attempt, recently made by Allan Hazlett in a paper entitled "The myth of factive verbs" (Hazlett 2010).

Despite its title, Hazlett's paper does not purport to show that none of the verbs that are commonly accepted by philosophers and linguists as factive is, in fact, factive (that is, is such that the unembedded sentence of which it is the main verb cannot be true unless the embedded sentence that functions as that verb's syntactic complement is also true). Indeed, toward the end of his paper, Hazlett asserts, speaking of several kinds of predicate expressions "that are often called 'factive'" (Hazlett 2010: 519), that what his paper contends "is consistent with some or all of these expressions being 'factive'" (Hazlett 2010: 519–520). The chief claim to which Hazlett definitely wants to commit himself is that there is *one* expression (the verb "know"), in *one* type of syntactic environment (the one in which the expression's syntactic complement is a *that*-clause rather than a *wh*-clause), which should *not* be taken to be factive, even though it is generally taken to *be* factive. "I have made a case for denying", he says in summarising his paper, "that an utterance of 'S knows p' is true only if p is true, i.e. that 'knows' is factive" (Hazlett 2010: 519). Although Hazlett's goal is thus much more limited than his paper's title suggests, it seems to me very doubtful, as I propose to explain in this essay, that even that limited goal has been attained.

The virtually universal acceptance of the thesis that Hazlett wants to deny— that is, of the thesis that "S knows that p" entails p—is ultimately based on the observation that, for innumerable choices of p, the conjunction of "S knows that p" with the negation of p is clearly unacceptable, an observation whose best explanation is taken to be that the conjunction expresses the obvious contradiction

that it would express if the hypothesised entailment relation obtained. Thus, utterances of (1), (3), (5) and (7) are clearly unacceptable—in sharp contrast to, say, utterances of (2), (4), (6) and (8), which are clearly acceptable:[1]

(1) * John knows that someone is watching him, but no one is.
(2)   John believes that someone is watching him, but no one is.

(3) * John knows that the book is in the box, but it isn't.
(4)   John believes that the book is in the box, but it isn't.

(5) * John knows that triangles have no sides, but they do.
(6)   John believes that triangles have no sides, but they do.

(7) * John knows that bachelors have wives, but they don't.
(8)   John believes that bachelors have wives, but they don't.

And these sharp acceptability contrasts, it is generally supposed, are best explained by assuming that (1), (3), (5) and (7) are contradictory, whereas (2), (4), (6) and (8) are not—in other words, that "S knows that p" entails p, whereas "S believes that p" does not entail p.

An especially important, though not often emphasised, reason in favour of the hypothesis that "S knows that p" entails p is the one illustrated by the contrast between (5) and (7), on the one hand, and (6) and (8), on the other—the reason, namely, that, unless one accepted that hypothesis, one would have to accept that necessary falsehoods can be objects of knowledge (to accept, in other words, that it is possible for someone to *know* that something is the case even though it is *impossible* for that thing to be the case). Few would suppose that accepting this would be anything else than absurd (unlike accepting that it is possible for someone to *believe* that something is the case even though it is impossible for that thing to be the case), and therefore few would suppose that attempts to show that there is no entailment from "S knows that p" to p—and hence no sense in which "know" is factive—could be successful.

Hazlett (who fails to consider cases where the complement of "know" expresses a necessary falsehood, even though such cases would be crucially relevant to evaluating any proposal, such as the one he puts forward, about the meaning of "know") does not dispute the fact that there exist conjunctions of "S

---

1 Asterisks are used to mark unacceptability, rather than ungrammaticality, throughout this chapter.

knows that *p*" with the negation of *p* that clearly *are* unacceptable. He claims, however, that there *also* exist conjunctions of "S knows that *p*" with the negation of *p* that are acceptable, and wants to conclude, on the basis of that evidence, that "S knows that *p*" does not entail *p*, and hence that "know" is not factive. Specifically, Hazlett cites exactly *two* examples of acceptable utterances, reproduced as (9) and (10) below, where the truth of attributed pieces of knowledge is implicitly denied, and claims that, since "the best theory of knowledge attributions" (Hazlett 2010: 501) that would be capable of explaining their acceptability is the one that would explain it by supposing that these utterances "could be true" (Hazlett 2010: 501)—in other words, that they are not contradictory—it follows that "S knows that *p*" does not entail *p*:

(9)   Everyone knew that stress caused ulcers, before two Australian doctors in the early 1980s proved that ulcers are actually caused by bacterial infection.

(10)   He figures anything big enough to sink the ship they're going to see in time to turn. But the ship's too big, with too small a rudder...it can't corner worth shit. Everything he knows is wrong.

Though neither of these examples—which are claimed to be authentic—has the form "S knows that *p*, but ~*p*", which would be relevant to Hazlett's stated argumentative goals, I will be assuming for the sake of argument—and probably too charitably, as far as the second example is concerned—that acceptable transformations of them into the relevant form could be effected. It is worth noting, incidentally, that the first of these examples could not be taken by Hazlett as evidence that the verb "know" is not factive *unless* it were at the same time taken by him as evidence that the verb "prove" *is* factive—i.e. that "S proved that *p*" does entail *p*. And similarly—to use two examples that Hazlett does not give, but might have given, in the same direction—the acceptability of (11) and (12),

(11)   We all knew that Mary was Tom's wife, until one of us discovered today that she was never actually married.

(12)   We all knew that John was a devout Christian, until he revealed to us today that he has always been an atheist.

could not be taken by Hazlett as evidence that "know" is not factive *unless* it were at the same time taken by him as evidence that "discover" and "reveal" *are* factive—i.e., that "S discovered that *p*" does entail *p* and that "S revealed that *p*"

does entail *p*. This shows—among other things, one of which will be noted below—that Hazlett's purported argument to the effect that "know" is not factive not only "is consistent with" but actually *requires* the view that several other verbs *are* factive—contrary, of course, to what both the title and the text of his paper would encourage one to suppose.[2]

2 Against the no-entailment view, Part I

Now, someone who, like Hazlett, is willing to regard the acceptability of certain utterances of the form "S knows that *p*, but ~*p*" as evidence that what these utterances express is not contradictory, would be expected to be equally willing to regard the *un*acceptability of certain other utterances of the form "S knows that *p*, but ~*p*" as evidence that what these *other* utterances express *is* contradictory. And if such a person wanted to reconcile these two pieces of evidence, he would be expected to surmise either that the word "know" is *polysemous*—and in particular, that it is ambiguous between a factive and a non-factive sense—or that it is a word that is semantically *underspecified*—i.e., that there is nothing that *is* its context-independent full meaning, but rather that, context-independently, its meaning is essentially incomplete, and capable of various kinds of contextually guided completions (some of which render some of its occurrences factive and some of which render some other of its occurrences non-factive).

   Hazlett, however, does not proceed in any of these ways: Although he unreservedly regards the acceptability of certain utterances of the form "S knows that *p*, but ~*p*" as evidence that these utterances do not express contradictions, he simply refuses to regard the *un*acceptability of certain other utterances of the *same* form as evidence that those other utterances do express contradictions. Hazlett's only apparent motive for adopting this idiosyncratic methodology is that, were he not to adopt it, he would be forced to conclude that "know" is ambiguous, and in particular polysemous, a conclusion that he is anxious to avoid because he is determined to follow Grice's recommendation, popularly known as the 'Modified Ockham's Razor', that "senses are not to be multiplied beyond necessity" (Grice 1989: 47). "Following Grice", he says, "I take the positing of polysemy as a vice, ceteris paribus, for a linguistic theory" (Hazlett 2010: 503). (It should be noticed that neither this claim nor the Gricean recommendation itself

---

2 Note that the title's point is re-affirmed elsewhere in Hazlett's text; for example, in his statement, "I've called factive verbs a myth, and I mean something by that" (Hazlett 2010: 500).

takes into account the semantic under-specification option. It should also be noticed that, even on the assumption that compliance with the Gricean recommendation would be compulsory in the present context, it could be achieved not only by taking the acceptability data as evidence of non-contradictoriness while refusing to take the unacceptability data as evidence of contradictoriness, but, alternatively, by taking the unacceptability data as evidence of contradictoriness while refusing to take the acceptability data as evidence of non-contradictoriness. In fact, this alternative approach, which is considerably less radical than Hazlett's, since it satisfies the Gricean recommendation *without* abandoning the virtually universally held view that "S knows that *p*" entails *p*, is precisely the one that has been advocated, for the sort of data that Hazlett considers, by Kvanvig (2006: 89–92). It is therefore puzzling that Hazlett nowhere gives reasons for preferring the more radical approach to the less radical one, and does not even mention, let alone discuss, Kvanvig's alternative Gricean approach to the matter.)

Now, there are several theorists who are decidedly less enthusiastic than Hazlett (or Kvanvig) is about following the Gricean recommendation, presumably because they consider that, since ambiguity in general, and polysemy in particular, is a fundamental feature of natural languages that most clearly distinguishes them from formal languages, the proper response to its existence would be to try to systematically investigate its empirical conditions of emergence and maintenance rather than to imagine that it can be either eliminated or restricted by means of appeals to theoretical economy. As Millikan (2005: 196) eloquently puts it,

> Ockham's razor employed to prohibit proliferation of semantic meanings can be as useless as it is for prohibiting the proliferation of living species.

But even those who are in general sympathetic to the Gricean recommendation would be unwilling to follow it in those cases in which doing so would require inventing idiosyncratic methodologies such as the one employed by Hazlett. Grice himself, at any rate, does *not* follow it under such circumstances. For example, Grice's theory of meaning famously relies on a fundamental distinction between two different *senses* of the word "mean", the so-called 'natural' sense, in which "X means that *p*" entails *p*, and the so-called 'non-natural' sense, in which "X means that *p*" does not entail *p*; and this is, of course, a distinction between a factive and a non-factive sense of "mean", as Grice explicitly recognises: "the non-natural cases of meaning", he points out, "are what we might call non-factive, whereas the natural cases are factive" (Grice 1989: 291). Furthermore, Grice's proposed grounds for claiming that "mean" is ambiguous between (at least) a

factive and a non-factive sense is that the posited ambiguity can explain the sharp contrast in acceptability between a sentence like (13) and a sentence like (14), by attributing the unacceptability of (13) to the contradictoriness of what it express, and the acceptability of (14) the non-contradictoriness of what it express:

(13)   *   These spots on your face mean that you have measles, but you don't have measles.

(14)   That sentence on your computer screen means that life is easy, but life is not easy.

Contrary to what Hazlett suggests, then, Grice does *not* deny that distinct senses of a word—for example, a factive and a non-factive sense of the word "mean"—can and must be posited when the alternative to positing them would be to invent idiosyncratic methodologies such as that of taking acceptability as evidence of non-contradictoriness while refusing to take unacceptability as evidence of contradictoriness. And since employing such an idiosyncratic methodology is exactly what Hazlett does, the antecedent plausibility of his approach is highly doubtful. Let us ignore, however, the general qualms that one might have about Hazlett's methodology and proceed to examine how exactly he proposes to justify his position with respect the particular case he is concerned with—that is, how exactly he proposes to account for the clear *un*acceptability of innumerable utterances of the form "S knows that *p*, but ~*p*", given that he rejects its standard explanation in terms of contradictoriness.

Hazlett's account is in two parts. The first part consists in the suggestion that, although a sentence of the form "S knows that *p*" does not semantically entail *p*, its utterance nevertheless *conversationally implicates* what is expressed by *p*. [Hazlett refers to the alleged conversational implicatures simply as "implications", but makes it clear that it is the Gricean notion of conversational implicature that he has in mind: unlike entailments, the 'implications' in question are supposed to concern the "pragmatics" (Hazlett 2010: 514) rather than the semantics of utterances of "S knows that *p*"; and they are supposed to be the results of pragmatic inferences that not only are explicitly referred to as "Gricean" (Hazlett 2010: 511, 520, 521) but do not make use of any structural resources that are not among those supplied by Grice's doctrine of conversational implicatures—in particular, the "principle of conversational cooperation" and its associated maxims of "Quality," "Quantity" and "Relation" (Hazlett 2010: 511–512).] The second—and, for present purposes, the crucial—part of Hazlett's account consists in the

suggestion that the supposed fact that "S knows that *p*" conversationally implicates *p* "may explain" (Hazlett 2010: 506) why the conjunction of "S knows that *p*" with the negation of *p* is unacceptable.

It requires but little reflection to see that, within the Gricean framework that Hazlett claims he assumes, his suggested explanation of the unacceptability of utterances of the form "S knows that *p*, but ~*p*" is incoherent. According to Grice, a proposition communicated by the utterance of a sentence is a conversational implicature of that sentence *only if* it is *cancellable*, that is, only if "to the form of words the utterance of which putatively implicates that *p*, it is admissible to add *but not p*" (Grice 1989: 44). (Notice that reliance on the cancellability requirement is not optional for those employing the notion of conversational implicature: since there would be little point to the notion of conversational implicature if conversational implicatures could not be effectively distinguished from entailments, and since the primary evidence held to ensure the possibility of such a distinction is that conversational implicatures are cancellable whereas entailments are noncancellable, not accepting cancellability as a requirement on conversational implicature would remove one's primary basis for supposing that conversational implicatures, as distinct from entailments, exist at all.[3]) Given, then, that cancellability is a necessary feature of conversational implicature and that, hence, "all conversational implicatures are cancellable" (Grice 1989: 44), it follows that no utterance of "S knows that *p*" could conversationally implicate *p* unless it were *admissible* to conjoin it with the negation of *p*—unless, that is, it were *acceptable*, rather than unacceptable, to expand that utterance into "S knows that *p*, but ~*p*." But if what *p* expresses cannot be a conversational implicature of an utterance of "S knows that *p*" unless that utterance can be *acceptably* expanded into "S knows that *p*, but ~*p*", it is incoherent to suppose, as Hazlett does, that the *un*acceptability of utterances of "S knows that *p*, but ~*p*" can be 'explained' by the hypothesis that what *p* expresses *is* a conversational implicature of "S knows that *p*." On the contrary: compatibility with the thesis that what *p* expresses is *not* a conversational implicature of "S knows that *p*" would be, on Grice's view, a necessary condition on any adequate explanation of the unacceptability of utterances of the form "S knows that *p*, but ~*p*."

It will of course be noticed that the hypothesis that Hazlett was trying to avoid by appealing to conversational implicatures—namely, the hypothesis that "S

---

3 Weiner's (2006) recent discussion of cancellability, in addition to other limitations pointed out by Blome-Tillmann (2008) and by Borge (2009), does not clearly acknowledge this fundamental point.

knows that *p*" semantically entails *p*—clearly fulfils the above-mentioned condition, since, just as cancellability is a necessary feature of conversational implicature, *non*-cancellability is, for Grice as for everyone else, a necessary feature of entailment—it is not surprising, in view of this fact, that, in an incidental discussion of the topic of the analysis of knowledge, Grice (1989: 53) explicitly endorses the view that Hazlett wants to deny, namely, the view that *p* is a semantic *entailment* of "S knows that *p*". And since the semantic explanation of the unacceptability of utterances of the form "S knows that *p*, but ~*p*" in terms of the hypothesis that "S knows that *p*" entails *p* is the only coherent explanation currently available, it is clearly the one that should be retained, especially since the hypothesis that affords that explanation presents independent advantages, such as the advantage of ensuring that necessary falsehoods cannot be objects of knowledge.[4]

3 Against the no-entailment view, Part II

Given that the semantic explanation of the unacceptability of utterances of "S knows that *p*, but ~*p*" in terms of their contradictoriness is the one to be retained, the conclusion that there is at least one sense of "know" in which that verb is factive cannot be avoided, contrary to what Hazlett was primarily aiming to establish. And if, in addition to accepting that conclusion, one wished to preserve Hazlett's claim that, in those relatively rare cases in which an utterance of "S knows that *p*, but ~*p*" is acceptable, its acceptability is due to its non-contradictoriness, the only available option would be to suppose, again contrary to Hazlett, that "know" has two distinct senses, a factive and a non-factive one.[5] Let me complete my case, then, by presenting some independent evidence suggesting that "know" does indeed have such distinct senses, and in particular that, in the two cases in which, according to Hazlett, the conjunction of "S knows that *p*" with the negation of *p* is acceptable, the verb "know" should be taken to have a non-factive sense distinct from the factive sense that it has in most other contexts.

The first of the two examples given by Hazlett [the one reproduced above as (9)], as well as the similar examples in (11) and (12) that I have offered in the same

---

4 For some independent problems facing Hazlett's purported derivation of *p* as a conversational implicature of "S knows that *p*", see the present paper's Appendix.
5 Hazlett obliquely admits the relative rarity of acceptable conjunctions of "S knows that *p*" with the negation of *p*, when he claims, speaking of uses of "S knows that *p*" whose conjunction with the negation of *p* are *un*acceptable, that they are "the most common uses of "knows" we have" (Hazlett 2010: 514).

direction, purport to attribute to a subject, as a state of knowledge, an epistemic state *that is at the same time alleged to have been based on evidence that has turned out to be unreliable* (recall that the examples are uninterpretable in the way intended by Hazlett *unless* the verbs "prove," "discover" and "reveal" are taken to *be* factive; and note that that assumption entails that the epistemic states that, in these sentences, are attributed to subjects as states of knowledge have been based on evidence that has turned out to be unreliable). But now notice that if the verb "know" did not have but a *single* sense, allowing it to apply to an epistemic state independently of whether or not the content of that state is evidentially undermined, sentences like (15) or (17) would be contradictory, just as, for example, sentences like (16) or (18) are:

(15)   If the President knew this just on the basis of unreliable evidence, he didn't really know it.

(16)   * If the President believed this just on the basis of unreliable evidence, he didn't really believe it.

(17)   If you know this simply on the basis of what these ignoramuses have been telling you, you don't know it at all.

(18)   * If you believe this simply on the basis of what these ignoramuses have been telling you, you don't believe it at all.

The fact of the matter, however, is that neither (15) nor (17) are contradictory, even though both (16) and (18) are. And that can only be because, in each of (15) and (17), the initial and final occurrences of the verb "know" have *different* senses, one of which allows it to apply to epistemic states whose contents have been evidentially undermined and the other of which prevents it from applying to epistemic states whose contents have been evidentially undermined. But these are surely among the contrastive features that occurrences of "know" would be expected to manifest if that verb had *distinct* factive and non-factive senses. And since it is the non-factive sense that "know" must evidently be supposed to have in the first of the two examples offered by Hazlett, what the acceptability (and non-contradictoriness) of that example must be taken to show is not that there is *no* sense in which "know" is factive but simply that there is *a* sense in which it is non-factive.

     Similar considerations apply, even more straightforwardly, to the second example offered by Hazlett and reproduced above as (10), where the acceptability (and hence, for Hazlett, the non-contradictoriness) of an utterance like "Every-

thing he knew was wrong" is supposed to show (on the assumption, one presumes, that "wrong" can only here mean "false") that there is no sense in which "know" is factive. The problem, again, is that if "know" did not have distinct factive and non-factive senses, (19) would be blatantly contradictory, just as (20) is:

(19)   If absolutely everything he knew was wrong, then he knew absolutely nothing.

(20) * If absolutely everything he believed was wrong, then he believed absolutely nothing

However, although (20) is contradictory, (19) isn't. And this can only be because, in (19), the initial and final occurrences of the verb "know" have *distinct* senses, of which the one is non-factive and the other factive. But if "know" has distinct factive and non-factive senses, what the acceptability and non-contradictoriness of (10) constitutes evidence for is simply the existence of the non-factive sense, and not, as Hazlett imagines, the non-existence of the factive sense.

The fact that "know", like several other verbs (including, as Grice had noticed, the verb "mean", or, as you yourself may have noticed, the verb "forget" in the title of the present essay), has distinct factive and non-factive senses is surely an interesting linguistic fact, and it may, or may not, have interesting philosophical implications.[6] Let me conclude by pointing out, however, that it does *not* have the epistemological implications that Hazlett would seem disposed to believe that it would have.

Having denied (without adequate evidence, as I have been arguing) that "know" has distinct factive and non-factive senses, and having opined that (contrary to what much current research suggests) epistemological analyses of knowledge should not be sensitive to linguistic evidence regarding the semantics of "know", Hazlett urges us to "note well" (Hazlett 2010: 519) that, even if epistemological analyses of knowledge *were* constrained to be sensitive to linguistic evidence regarding the semantics of "know", and even if "know" *did* have distinct factive and non-factive senses, the plurality of its senses would diminish the linguistic plausibility of "the traditional analysis of knowledge" (Hazlett 2010: 519)—that is, of the analysis that requires, among other things, that "S knows that *p*" entail *p*—for the following supposed reason:

---

6 See Stjernberg (2009) for some discussion of this issue.

> If "knows" has a plurality of meanings, then from a linguistic point of view there's nothing that makes the traditional [analysis of knowledge] *the* definition, other than epistemological tradition! (Hazlett 2010: 519)

Now, it is analytic that if a word has more than one meaning, and meanings are specifiable in terms of definitions, then to each distinct meaning of the word there corresponds a distinct definition of the word, and none among these definitions can be singled out as *the* definition of the word. But from the trivial fact that to each distinct meaning of a word there corresponds (assuming that meanings are given by definitions) a distinct definition of that word, it certainly does *not* follow that a field of study that takes as its subject matter what, in *one* of its meanings, the word denotes, suffers from a defect of linguistic implausibility simply because, in *another* one of its meanings, the word denotes other things. For example, truth is the subject matter of logic in one established sense of the word "true", and the fact that this word has another established sense in which it means "authentic" (as in "true diamonds" or "true aristocrats") neither augments nor diminishes the linguistic plausibility of the logician's choice of subject matter: it is simply linguistically irrelevant to that choice. Similarly, factive knowledge is the chosen subject matter of epistemology, and the fact that, besides its factive sense, the word "know" has another sense that is not factive neither augments nor diminishes the linguistic plausibility of the epistemologist's choice of subject matter: it is simply linguistically irrelevant to that choice. Of course, someone might wish to establish a new field, distinct for epistemology, dedicated to speculations about whatever it is that non-factive uses of "know" are supposed to denote (somewhat like Heidegger, who wanted to establish a new field, distinct from logic, dedicated to speculations about truth as authenticity). But I, for one, cannot see what the interest of establishing such a field might be; and, in view of the fact that the putative field's constitutive commitments would include the commitment to represent every falsehood, including every *necessary* falsehood, as a potential piece of 'knowledge', I wouldn't be optimistic about its prospects.

Appendix: Further adventures in Griceland

As explained in the text, Hazlett's hypothesis that utterances of "S knows that *p*" conversationally implicate, rather than entail, *p*, is precluded by the cancellability requirement; it may be interesting to note, however, that that hypothesis suffers from important further problems as well.

As is well known, the conversational implicatures of an utterance of a sentence are calculated, according to Grice, on the basis of what, in uttering that sentence, a speaker literally *says*. Hazlett, however, claims that when, in uttering a sentence of the form "S knows that *p*", a speaker conversationally implicates *p*, the alleged implicature is calculated simply on the basis of a certain thing that, in Hazlett's view, these sentences *entail*. Specifically, Hazlett hypothesises that sentences of the form "S knows that *p*" entail corresponding sentences of the form "S has the epistemically warranted belief that *p*"; he then claims that, since "it is a conceptual truth that epistemically warranted beliefs tend to be true" (Hazlett 2010: 508), and since, given this allegedly conceptually guaranteed 'tendency', a cooperative speaker who ascribes an epistemically warranted belief to a subject S will be taken to conversationally implicate, unless he provides explicit indications to the contrary, that the belief he ascribes to S as epistemically warranted is true, utterances of "S knows that *p*" come to conversationally implicate *p*.

There are crucial elements of this account that one might wish to question. Thus, one might argue that, contrary to what Hazlett assumes, sentences of the form "S knows that *p*" do *not* entail corresponding sentences of the form "S has the epistemically warranted belief that *p*", since, for example, (a) and (b) could be true even though (a′) and (b′) could be false (assuming that they would not be nonsensical):

(a)   John knows that he is angry about what Ann just did.
(a′)   John has the epistemically warranted belief that he is angry about Ann just did.
(b)   John knows that he has a headache.
(b′)   John has the epistemically warranted belief that he has a headache.

But even assuming that Hazlett's hypothesised entailment relation obtains, his account would face the problem that it would require acceptance of a clearly false thesis regarding conversational implicature. To see this, notice that Hazlett professes not to know anything about what "S knows that *p*" means *except* that it does not entail *p* and that it does entail "S has the epistemically warranted belief that *p*"—in particular, he is explicitly agnostic as to whether his epistemic warrant condition is "sufficient for the truth of 'S knows *p*'" (Hazlett 2010: 509), and he is equally agnostic as to "whether there are additional necessary conditions on the truth of 'S knows *p*'" (Hazlett 2010: 509). But anyone who is to that extent non-committal about what the truth conditions of "S knows that *p*" are—and so, about what, in the Gricean sense, utterers of sentences of that form literally *say*—

would hardly be entitled to make hypotheses as to what such utterers conversationally implicate (after all, implicature-generating inferences are supposed to be inferences that derive, *on the basis of a full specification of what speakers literally say*, what these speakers mean without literally saying). To justify his procedure, then, Hazlett would have to specify a special principle whose truth would entitle him to make hypotheses as to what utterers of "S knows that *p*" conversationally implicate, even in the absence of full information as to what they literally say. And, as far as I can see, the only principle that, if true, would provide him with the requisite entitlement, would be the following:

(HAZ)   If a sentence S_1 entails a sentence S_2, then whatever can be conversationally implicated by utterances of S_2 can also be conversationally implicated by utterances of S_1.

The principle stated in (HAZ), however, is certainly false. For example, an utterance of (i) can conversationally implicate (ii), but no utterance of (iii) can conversationally implicate (ii), even though (iii) entails (i):

(i)   It's cold in here.
(ii)   I want you to close the windows.
(iii)   It's cold in here, and all the windows are closed.

Similarly, an utterance of (iv) can conversationally implicate (v), but no utterance of (vi) can conversationally implicate (v), even though (vi) entails (iv):

(iv)   John is nervous sometimes.
(v)   It is not the case that John is always nervous.
(vi)   John is always nervous.

If, however, the principle stated in (HAZ) is false, then the supposed fact that "S knows that *p*" entails "S has the epistemically warranted belief that *p*", together with the supposed fact that utterances of "S has the epistemically warranted belief that *p*" conversationally implicate *p*, does *not* authorise the conclusion, which Hazlett was interested in drawing, that utterances of "S knows that *p*" *themselves* conversationally implicate *p*. In short, Hazlett's purported derivation of the truth of *p* as a pragmatic implicature, rather than as a semantic entailment, of utterances of "S knows that *p*" would be unsuccessful even if it were not already disqualified by the cancellability requirement.

Chapter 24
Three problems for the knowledge rule of assertion

1 Introduction

Compactly stated, Timothy Williamson's theory of assertion is the following (cf. Williamson 1996; 2000: 238–269; 2009: 303ff, 341ff):

> Assertion is the unique speech act *F* whose unique constitutive rule is the knowledge rule:
>> (Knowledge Rule) One must: *F* that *p* only if one knows that *p*.

My purpose in what follows is to argue that Williamson's main linguistic argument for the claim that the speech act of assertion is governed by the above-mentioned Knowledge Rule is not compelling.

Williamson's main linguistic argument for the claim that assertion is governed by the Knowledge Rule is an abductive argument to the effect that it is only by assuming that assertion is so governed that one can explain why sentences of the form *(X)* are unassertable—that is, could not without oddity be asserted—even though what they would assert, if they could be asserted without oddity, might be true:

(X)   A and I do not know that A.

Williamson's proposed explanation of the unassertability of sentences of the form *(X)* is the following:

> What is wrong [with sentences of the form 'A and I do not know that A'] can easily be understood on the hypothesis that only knowledge warrants assertion. For then to have warrant to assert the conjunction 'A and I do not know A' is to know that A and one does not know A. But one cannot know that A and one does not know A. One knows the conjunction only if one knows each conjunct, and therefore knows that A (the first conjunct); yet one knows the conjunction only if it is true, so only if each conjunct is true, so only if one does not know that A (the second conjunct); thus the assumption that one

knows the conjunction that A and one does not know that A yields a contradiction. Given that only knowledge warrants assertion, one therefore cannot have warrant to assert 'A and I do not know that A'. (Williamson 2000: 253; cf. Williamson 1996: 506)

To recapitulate: Williamson's chosen explanandum is the linguistic fact that sentences of the form *(X)* are unassertable, in the sense that they could not without oddity be asserted; his proposed explanation of this linguistic fact is that, on the assumption that the act of assertion is governed by the Knowledge Rule, the unassertability of these sentences would follow from the fact that it would be logically impossible for their assertion to be warranted; and the fact that the Rule, if it were assumed to be governing the act of assertion, would make that explanation possible constitutes, in Williamson's view, a decisive reason for thinking that it does govern the act of assertion.

I will present three problems for Williamson's argument, of which the first two show that, even assuming that Williamson's explanandum has been properly circumscribed, his explanation would not be correct, and the third shows that his explanandum has not been properly circumscribed.[1]

2 The first two problems

The first two problems arise from cases where a sentence of the form *(X)* is unassertable, but its unassertability cannot be explained by Williamson's account, since that account is insensitive to the crucially relevant *first-personal* character of the sentences whose unassertability it sets out to explain.

The case of the ignorant speaker

Suppose that Helen Wilson is amnesiac about her name, and that, incorrectly believing that she is someone *other* than Helen Wilson, she utters the sentences in (1) and (2):

(1)   John is hiding and Helen Wilson doesn't know that he is hiding.
(2)   It is five o'clock and Helen Wilson doesn't know that it is five o'clock.

---

1 Asterisks are used to mark unassertability throughout this chapter.

Sentences (1) and (2) are certainly *assertable* by Helen in the context just described. However, sentences (3) and (4)—which, were they to be uttered by Helen in the same context, would have exactly the same truth conditions that (1) and (2), respectively, have—are certainly *un*assertable ones:

(3)   ⋆   John is hiding and I don't know that he is hiding.
(4)   ⋆   It is five o'clock and I don't know that it is five o'clock.

The contrast between (1) and (2), on the one hand, and (3) and (4), on the other, shows that the reason why a sentence of the form *(X)* is unassertable cannot be the reason adduced by Williamson. According to Williamson, what makes a sentence of the form 'A and I don't know that A' unassertable is that, applied to it, the Knowledge Rule entails that if one is to assert such a sentence one must "know that A and one does not know A", which is a requirement that it is logically impossible to satisfy ("one cannot know A and one does not know A"); the hypothesis, then, that assertion is subject to the Knowledge Rule explains why such sentences are unassertable. However, if that were the reason why (3) and (4) are unassertable by Helen, (1) and (2) should *also* be unassertable by her. For, (1) and (2) have exactly the same truth conditions that (3) and (4), as uttered by Helen, would, respectively, have. So, if the unassertability of (3) and (4) by Helen were due to the fact that, given the Knowledge Rule, her assertion of them would commit her to satisfying logically unsatisfiable requirements, the same Knowledge Rule, applied to her utterances of (1) and (2), would commit her to satisfying the same logically unsatisfiable requirements, and so should prevent (1) and (2) from being assertable by her. But since (1) and (2) *are* assertable by her, it follows that it is not any requirements entailed by the Knowledge Rule that explain why (3) and (4) are *not* assertable by her. And since it was precisely its supposed capacity to explain the unassertability of sentences like (3) and (4) that justified, according to Williamson, the positing of the Knowledge Rule, it follows that Williamson's argument fails to provide reasons for positing the rule in question.

The case of the ignorant addressees

It might be thought that the sort of problem just noted can only be raised by reference to the atypical kind of case where, because of amnesia, a speaker does not know what his or her name is. In fact, however, exactly the same sort of problem can be raised by reference to the far from atypical type of case where an *addressee* happens not to know what the name of a speaker addressing him or her is. Suppose that Timothy Williamson is having a conversation with a group of people who are ignorant of the fact, and are known by Williamson to be ignorant of the

fact, that their interlocutor's name is "Timothy Williamson". Suppose further that these people are asking Williamson questions like "Is the Bodleian Library closed? And if it is, does Timothy Williamson know that it is?" or "Is the Covered Market open? And if it is, does Timothy Williamson know that it is?" Finally, suppose that, for some important reasons of his own, Williamson wants these interlocutors to *remain* in their state of ignorance regarding his name (perhaps he has excellent grounds for thinking that they would seriously harm him if they were to know what they presently ignore). To safeguard his vital interests, Williamson might then choose to utter, in a confident and reassuring tone of voice, the following responses to the questions addressed to him by his interlocutors:

(5)   The Bodleian Library is closed and Timothy Williamson doesn't know that it is.

(6)   The Covered Market is open and Timothy Williamson doesn't know that it is.

Sentences (5) and (6) are certainly *assertable* by Williamson, in the context just described. However, sentences (7) and (8)—which, were they to be uttered by Williamson in the same context, would have exactly the same truth conditions that sentences (5) and (6), respectively, have—, are clearly *un*assertable ones:

(7)   ⋆   The Bodleian Library is closed and I don't know that it is.

(8)   ⋆   The Covered Market is open and I don't know that it is.

Just as in the previous type of case, the contrast between (5) and (6), on the one hand, and (7) and (8), on the other, shows that Williamson's account of what explains the unassertability of sentences of the form *(X)* cannot be right. According to that account, the reason why a sentence of the form 'A and I don't know that A' is unassertable is that, applied to it, the Knowledge Rule entails that if one is to assert such a sentence one must "know that A and one does not know A", which is a logically unsatisfiable requirement ("one cannot know A and one does not know A"); the hypothesis, then, that assertion is subject to the Rule explains why the sentences in question are unassertable. However, if that were the reason why (7) and (8) are unassertable by Williamson, (5) and (6) should *also* be unassertable by him. For, (5) and (6) have exactly the same truth conditions that (7) and (8), as uttered by Williamson, would, respectively, have. So, if the unassertability of (7) and (8) by Williamson were due to the fact that, given the Knowledge Rule, his assertion of them would commit him to satisfying logically unsatisfiable requirements, the same Knowledge Rule, applied to his utterances of (5) and (6),

would commit him to satisfying the same logically unsatisfiable requirements, and so should prevent (5) and (6) from being assertable by him. But since (5) and (6) *are* assertable by him, it follows that it is not any requirements entailed by the Knowledge Rule that explain why (7) and (8) are *not* assertable by him. And since it was precisely its supposed capacity to explain the unassertability of sentences like (7) and (8) that justified, according to Williamson, the positing of the Knowledge Rule, it follows that Williamson's argument does not provide reasons for positing the rule in question.

3 The third problem

Since, as the above problems make clear, the unassertability of sentences of the form *(X)* crucially depends on the fact, which is left out of consideration in Williamson's explanation, that the subject of the knowledge-denials they contain is first-personal, Williamson's abductive argument could not be saved unless the Knowledge Rule were somehow changed so as to become sensitive to that fact. I cannot see any non-arbitrary way in which this could be done, but I will not pursue this matter further, since, as the problem to which I now turn shows, the explanatory project through which Williamson was aiming to justify the postulation of the Knowledge Rule would be bound to fail in any case. The source of this further problem is that, contrary to what that project explicitly assumes, there exist not only unassertable but also *assertable* sentences of the *first-personal* form 'A and I don't know that A'.

Suppose a person accepts a form of determinism (call it 'mysterian determinism', to give it a name) according to which, although there is no such thing as free will, we are biologically engineered so as to be unable to ever *know* that there is no such thing as free will. It would not be strange for a person who accepts this doctrine to produce, among others, *modus ponens* arguments like the following:

(9)   If I am not omniscient, my future actions are predetermined and I don't
         know that they are.
       I am not omniscient.
       Therefore, my future actions are predetermined and I don't know that
         they are.

It is clear that, *in the context provided by its premises*, the conclusion of this inference is not at all unassertable. However, the conclusion of the inference is an in-

stance of schema *(X)*, and therefore contradicts the basic assumption of Williamson's explanatory project that there are *no* assertable instances of that schema. Notice that what Williamson was aiming to explain by positing the Knowledge Rule was that "Something is wrong with *any* assertion of the form '[A] and I do not know that [A],' even though such assertions would often be true if made" (Williamson 1996: 506; emphasis added). Not surprisingly, then, Williamson's explanatory strategy breaks down when confronted with a sentence of the form *(X)* which, contrary to what the strategy assumes, *is* assertable. For, the fact that such a sentence is assertable can then only be taken to mean either that there is no Knowledge Rule at all governing assertion or that, even supposing that assertion is governed by such a rule, the fact that it is so governed can do nothing to prevent the sentence's assertability, and therefore is of no use in an abductive argument aiming to justify the rule's postulation.

It might be claimed that assertable instances of schema *(X)* are confined to contexts of philosophical argument of the sort invoked by the inference in (9), and should, for that reason, be set aside for special treatment. It is doubtful that the second part of this claim would be accepted as methodologically sound even if the first were true, but what is even more important to realise for present purposes is that the first part is simply not true: assertable instances of schema *(X)* can be encountered in perfectly ordinary contexts as well. For example, an ordinary speaker might very well produce the following instance of *modus ponens*:

(10)   If I am gullible, people lie to me and I don't know that they do.
        I am gullible.
        Therefore, people lie to me and I don't know that they do.

There is certainly nothing wrong with asserting the conclusion of this inference, *given the context provided by its premises*. The conclusion, however, is an instance of schema *(X)*, and so its assertability falsifies Williamson's assumption that "Something is wrong with *any* assertion of the form '[A] and I do not know that [A]', even though such assertions would often be true if made". With that assumption gone, one would be entitled to conclude either that there is no Knowledge Rule at all that governs assertion or that, even if such a rule were on independent grounds supposed to govern assertion, its supposed existence would simply be *irrelevant* to understanding why some instances of schema *(X)* are unassertable and some others aren't. And, of course, since the assumption that the rule was *not* explanatorily irrelevant was the principal reason purportedly supporting the belief in its existence, Williamson's abductive argument for its existence cannot be supposed to have been a successful one.

4 Conclusion

I have given three reasons for thinking that Williamson's main linguistic argument for the claim that assertion is governed by the Knowledge Rule is not compelling. In view of the fact that, without the Knowledge Rule, there would be nothing left to Williamson's account of assertion, that account itself, in so far as it relies on that argument, is not compelling either. Concerning the linguistic phenomenon that had prompted Williamson's argument, and which might legitimately be regarded as calling for an explanation, the outcome of the present discussion is that its adequate explanation, whatever it may finally turn out to be, should respect two conditions: First, it should be consistent with the fact that only *some*, and not *all*, instances of the schema 'A and I don't know that A' are unassertable. Second, it should be consistent with the fact that, when an utterance that is an instance of that schema *is* unassertable, replacing its first-personal referring term with a co-referential non-first-personal referring term may result in a truth-conditionally equivalent utterance that is *not* unassertable. In short, and perhaps not unexpectedly, the phenomenon is considerably more complex than initial appearances might suggest.

Chapter 25
Grammars as objects of knowledge: The availability of dispositionalism

1 Introduction

As is well known, Chomsky and several linguists following him believe that the rules and principles comprising the grammar of any given natural language are things that are *known* by speakers of that language—and that speakers possess that knowledge independently of any linguist's succeeding, or even trying, to formulate the rules and principles in question. Chomsky and his followers, however, do not believe that all reinterpretations of their professed belief are correct. In particular, they are strongly opposed to a reinterpretation of their professed belief that would appear to radically impoverish its content by adopting the—on at least one construal, characteristically Wittgensteinian—view that ascriptions of knowledge in a given domain are logically indistinguishable from ascriptions of dispositions to behave in certain ways in that domain (cf. Wittgenstein 1953, 1980). For, given that view, the claim that the rules and principles comprising the grammar of a natural language are things known by speakers of that language, can easily be taken to amount to nothing more than the claim that these speakers are disposed to exhibit certain patterns of verbal behaviour that happen to be correctly predictable (under appropriate idealisations) on the basis of those rules and principles, *without* those rules and principles being in any way represented in the minds or brains of the speakers. And this certainly appears to be an impoverishment of the idea—recurring with remarkable stability in Chomsky's works— that the sorts of rules and principles that he and his followers are proposing do not merely constitute appropriate bases for correct predictions about linguistic behaviour, but are also *causally* involved in the production of that behaviour (an involvement that would, of course, be impossible unless the rules and principles in question were somehow represented in the minds or brains of the subjects whose behaviour they are alleged to causally influence).

Chomsky's reason for rejecting the suggestion that a system of grammatical rules and principles, if accepted at all, might be construed merely as predictively reliable and not as causally efficacious is not an empirical one. For, as Chomsky is aware, no one has yet succeeded in locating the neural embodiments of any

specific grammatical rules and principles, and no one is, accordingly, in a position to empirically defend the claim that a particular system of rules and principles, as opposed to innumerable others with the same predictive power, is causally involved in the brain's control of linguistic behaviour. What is more, the very idea that *some* such system, whether or not it is at present neurologically identifiable, *must* be supposed to be involved in the brain's control of linguistic behaviour appears to progressively lose its original aura of inevitability with the advent and expansion of connectionist approaches to brain modelling (which, incidentally, may be seen as offering perspectives complementary to the Wittgensteinian ones; cf. Mills 1993); for, in many cases, these approaches have been successful in showing that the assumption that the brain controls a pattern of behaviour which is describable by means of a given set of rules or principles is consistent with the assumption that the brain does *not* refer to representations of *any* rules or principles in order to exercise its control; if, therefore, these approaches are correct even in some cases, they certainly block any *automatic* inference from the describability of a behavioural pattern by means of a rule system to the conclusion that the organisms exhibiting the pattern possess and consult neural representations of that rule system.

Chomsky, however, believes that, although he is not, given the present state of the neurosciences, in a position to demonstrate that a proposed system of grammatical rules and principles is indeed causally efficacious, he *is* in a position to justifiably assert, in advance of any empirical inquiry, that such a system's predictive utility does *not* exhaust its content. For, the claim that such a system's predictive utility does exhaust its content—a claim, it should be noted, that would also be embraced by many philosophers, like Quine (1970), who could hardly be described as typical Wittgensteinians—crucially depends, in Chomsky's opinion, on the Wittgensteinian idea that ascriptions of knowledge in a given domain are logically indistinguishable from ascriptions of dispositions to behave in certain ways in that domain. And, according to Chomsky, there is a decisive *a priori* argument against that idea, at least as far as the domain in question is the domain of language—there is, in other words, a decisive *a priori* argument to the effect that the concept of linguistic knowledge cannot be identified with the concept of a disposition to exhibit linguistic behaviour.

After describing how this *a priori* argument was formulated in Chomsky (1980), criticised in Kenny (1984), and defended against Kenny's critique in Chomsky (1988), I shall inquire in what follows whether Chomsky's defence is successful, reaching the result that it is not. My first conclusion will therefore be that, despite appearances, the dispositionalist position is *not* vulnerable to

Chomsky's critique. I will then examine whether the dispositionalist has the conceptual resources not only for countering Chomsky's attack but also for mounting an original attack of his own against the anti-dispositionalist position that Chomsky represents; and, finding that he does indeed have such resources, I shall derive my second, complementary conclusion, that it is Chomsky's position, rather than the dispositionalist's, that emerges as the really weak one in this particular debate.

2 The terms of the dispute

Chomsky's claim that a person's linguistic knowledge cannot be identified with that person's disposition to exhibit linguistic behaviour comes as a result of the following thought experiment:

> Imagine a person who knows English and suffers cerebral damage that does not affect the language centers at all but prevents their use in speech, comprehension, or let us suppose, even in thought. Suppose that the effects of the injury recede and with no further experience or exposure the person recovers the original capacity to use the language. In the intervening period, he had no capacity to speak or understand English, even in thought, though the mental (ultimately physical) structures that underlie that capacity were undamaged. Did the person know English during the intervening period? (Chomsky 1980: 51)

Chomsky's answer is that the protagonist of the above thought experiment did know English during the intervening period, and that, therefore, linguistic knowledge cannot be identified with the capacity to exhibit linguistic behaviour.

Kenny's reply is that Chomsky's argument cannot impugn the propriety of that identification, since one can legitimately maintain that the identification is correct while plausibly redescribing the imagined situation by invoking a distinction—whose independent desirability was pointed out by Kenny (1975) at an earlier time, and which was further elaborated in Kenny (1989)—between the existence or non-existence of a behavioural capacity and the existence or non-existence of the capacity *to exercise* that behavioural capacity. Considering Chomsky's question, "Did the person know English during the intervening period?", Kenny responds:

Well, we can say what we like as long as we know what we are doing: it is up to us to decide whether what is left is sufficient for us to call it "knowledge of English". Perhaps Chomsky is right that the more natural decision is to say that it is sufficient. Fine, then, let us say that the person knows English. But why should we not also say that the person retains the capacity to speak English? For extraneous reasons, he cannot use or exercise that capacity at the moment; but since, *ex hypothesi*, he is going to use it in future without any of the normal acquisition processes, is it not natural to say that he still holds on to it in the meantime? The concept of *capacity to use English* has exactly the same fuzzy edges as the concept of *knowledge of English* and Chomsky's example does nothing to separate the two concepts. (Kenny 1984: 141)

Chomsky's response to Kenny's reply is that it is ineffectual, since the distinction between the existence of a capacity and the existence of the capacity to exercise that capacity does not correspond to anything present in our ordinary concept of a capacity, and merely represents an arbitrary move designed to save the dispositional account. After repeating what he regards as the proper conclusions to draw from his thought experiment, Chomsky notes:

To avoid these conclusions, philosophers committed to the identification of knowledge and ability have been forced to conclude that [the person] who lost the ability to speak and understand [English] after brain injury in fact retained this ability, though he lost the ability to exercise it. We now have two concepts of ability, one referring to the ability that was retained and the other to the ability that was lost. The two concepts, however, are quite different. It is the second that corresponds to ability in the sense of normal usage; the first is just a new invented concept, designed to have all the properties of knowledge. Not surprisingly, we can now conclude that knowledge is ability, in this new invented sense of "ability" that is quite unrelated to its normal sense. Plainly nothing is achieved by these verbal maneuvers. We must conclude, rather, that the attempt to account for knowledge in terms of ability (disposition, skill, etc.) is misconceived from the start. (Chomsky 1988: 11–12)

Indeed, not only is Chomsky convinced that the Wittgensteinian account of knowledge in terms of ability is vitiated by being forced to adopt a conception of "ability" that diverges, according to him, from "normal usage", but he also contends that, by being forced to adopt such a divergent conception, it contradicts Wittgenstein's own much more general thesis that departures from "normal usage" should be always avoided because they are among the primary sources of

conceptual confusion. As Chomsky puts it in a later presentation of his response to Kenny's reply, "the Wittgensteinian construal of knowledge as a species of ability seems to be a paradigmatic example of the practice that Wittgenstein held to be a fundamental source of philosophical error" (Chomsky 1992: 104). Apparently, then, Wittgenteinians would have to concede not merely a local but a global defeat if Chomsky's suggestions could withstand scrutiny—which is one more reason for finding out whether in fact they could.

3 The structure of the dispute

Let Q stand for the person involved in Chomsky's thought experiment and *t* for the period between that person's injury and recovery. Chomsky's original argument can then be summarised by saying that, since statement (1a) below is true of Q at *t*, and since statement (1b) below should also be true of Q at *t if* the dispositional account of knowledge was correct, proponents of that account would be constrained by the canons of deductive reasoning to accept as also true of Q at *t* the statement in (1c), even though that statement is, in Chomsky's view, *not* true of Q at *t*.

(1)   a.   Q does not have the ability to use English
         b.   Q knows English if and only if Q has the ability to use English.
         c.   Therefore, Q does not know English.

In his reply, Kenny grants to Chomsky the right to assert that Q does know English at *t*, but insists that saying of Q that he cannot *exercise* the ability to use English at *t* (rather than that he does not *have* the ability to use English at *t*) is all that is required to truthfully represent Q's predicament. Once this is admitted, however, there is no way—Kenny suggests—in which a dispositionalist could be threatened by the inference that Chomsky is planning to impute to him. For, the inference in question will then have to take either the form in (2), which, though faithfully representing the dispositionalist position in premise (2b), is *not* deductively valid, or the form in (3), which, though deductively valid, does *not* faithfully represent the dispositionalist position in premise (3b):

(2)   a.   Q cannot exercise the ability to use English.
         b.   Q knows English if and only if Q has the ability to use English.
         c.   Therefore, Q does not know English.

(3)   a.   Q cannot exercise the ability to use English.

        b.   Q knows English if and only if Q can exercise the ability to use English.

        c.   Therefore, Q does not know English.

In his rejoinder, Chomsky correctly perceives that the only way to counter Kenny's objection is to deny that, as far as our ordinary concepts are concerned, there is a real distinction to be made between the idea of a person's *being able or unable to exercise an ability* and the idea of that person's *having or not having that ability*. He therefore thinks he can dispose of Kenny's objection by swiftly proceeding to that denial: there can be no possible difference, Chomsky contends, between the existence or non-existence of an ability and the existence or non-existence of the ability to exercise that ability; and since no such difference can possibly exist, Kenny is not entitled, Chomsky argues, to invoke it in order to defuse his objection to dispositionalist accounts of knowledge ascriptions. The question before us is, then, whether Chomsky is justified in his claim that statements of the form (4) are equivalent to corresponding statements of the form (5)—in other words, whether he is justified in his claim that every statement of the form (6) is necessarily true:

(4)   *x* has the ability to *f*

(5)   *x* can exercise the ability to *f*

(6)   *x* has the ability to *f* if and only if *x* can exercise the ability to *f*

4 A dispositionalist defence

In order to show that Chomsky's equivalence thesis does not in fact hold, it would be sufficient to show that there are circumstances where a statement of the form (4) would be true whereas a corresponding statement of the form (5) would be false. Showing this, however, is not at all difficult, since circumstances where a subject's inability to exercise a certain ability is consistent with that subject's continued possession of that ability are not at all rare.

     The abilities (or skills) to play football, to conduct an orchestra, to swim and to dance are certainly abilities that can truthfully be ascribed to persons. Now, each person to whom these or any other abilities are truthfully ascribed is, like any other person, an organism whose continued existence requires the regular immersion in states of sleep. During sleep, none of the abilities just mentioned can, of course, be exercised (and, during deep sleep, a host of many other, even

more common, abilities cannot be exercised either). From the fact, however, that, during sleep, an ability cannot be exercised, it hardly follows that it is *lost*. We would certainly not say that a football player, an orchestral conductor, a swimmer and a dancer lose their playing, conducting, swimming and dancing skills every time they go to sleep, and newly acquire them every time their sleeping sessions are over. Indeed, if these abilities were literally *lost* during each sleeping session, it would be a miracle that their bearers instantly reacquire them after each sleeping session, since these abilities cannot be acquired *at all* without considerable training (which is why they are commonly described as skills). The solution, of course, is to say that what the subjects in question lose when they fall asleep are *not* the above mentioned abilities but rather the ability *to exercise* those abilities—and, correlatively, that what they newly acquire when they wake up are *not* the above mentioned abilities but rather the ability *to exercise* those abilities. It is in order to conceptually accommodate perfectly ordinary situations like these that the distinction invoked by Kenny has been designed. And since it does succeed in accommodating them, its legitimacy cannot credibly be denied.

On the contrary, it is Chomsky's refusal to accept that distinction that cannot be legitimised, since it leads to palpably absurd redescriptions of the perfectly ordinary situations we have been considering. For, given that, according to Chomsky, an ability exists if and only if it can be exercised, Chomsky would have to say that an experienced football player and an experienced orchestral conductor completely *lose* they playing and conducting skills *every* time they go to sleep (since, during sleep, these skills cannot be exercised). And since a completely lost skill cannot reappear unless it is newly acquired, Chomsky would have to explain the disturbing fact that, once awake, the player and the conductor are immediately capable of manifesting their playing and conducting abilities, by saying that, though completely lost during sleep, these complex abilities are *instantly* reacquired by their bearers after *every* sleeping session (which, of course, amounts to saying that the ability to play football or to conduct an orchestra are, by turns, totally lost and totally regained during *each day* in the lives of their bearers). Now, "normal usage", which Chomsky appears willing to invoke on his behalf, may be notoriously tolerant, but even "normal usage" would not tolerate that ordinary lives are, as Chomsky's proposal implies, constant successions of miracles. Indeed, by providing the two *distinct* expression types '*x* has the ability to *f*' and '*x* can exercise the ability to *f*' without in any way *imposing* on us to treat them, in Chomsky's way, as necessarily co-extensive, "normal usage" performs two excellent services: first, it provides us with the opportunity of *inquiring* whether they are in fact co-extensive; and secondly, it provides us with the means, once we realise that they are *not* in fact co-extensive (once we realise, for

example, that the abilities we cannot exercise during sleep are not lost during each sleeping session and instantly reacquired after each sleeping session), of explaining why ordinary lives are *not* constant successions of miracles.

Our conclusion must be, then, that, since the distinction between a subject's possessing an ability and a subject's being able to exercise that ability is both real and vital in preventing the generation of manifest absurdities, Kenny had every right to invoke it in objecting to the anti-dispositionalist argument that Chomsky had built on the basis of his thought experiment; and that, consequently, Chomsky's attempt to save that anti-dispositionalist argument by denying the distinction's viability is unsuccessful. As a result, Kenny's original argument retains its full force, as well as the full range of its implications: Chomsky has *not* yet shown, contrary to what he and many of his followers may believe, that there is anything wrong with the proposal of identifying knowledge in the linguistic domain with the ability to exhibit certain behavioural patterns in that domain; and he has not, therefore, provided a decisive *a priori* reason for denying that a proposed system of grammatical rules and principles, if accepted at all, can be construed merely as a basis for accurate predictions (under appropriate idealisations) of certain behavioural regularities and not as a causal force responsible for these regularities.

5 A dispositionalist attack

Of course, the fact that Chomsky has failed to show that the dispositionalist position must be ruled out does not in itself determine whether his own anti-dispositionalist position is or is not ultimately defensible. It would be interesting to examine, therefore, whether the dispositionalists would be more successful in arguing, in their turn, that it is the anti-dispositionalist position that ought to be ruled out. It seems, in fact, that they could do precisely that, by exploiting a subtle, and no doubt unintended, concession to dispositionalism that is implicit in the way in which Chomsky's thought experiment is set up.

Recall that the subject of Chomsky's thought experiment is understood as a subject who *did behaviourally manifest* its knowledge of English *before* its injury and also as one who *did behaviourally manifest* its knowledge of English *after* its recovery. It is only after describing the subject in such terms that Chomsky goes on to raise the question whether it knew English "during the intervening period"—during the period, that is, in which it did not behaviourally manifest its knowledge—and argues that the affirmative reply to that question, which he takes to be correct, creates a problem for the dispositionalist. Now, as we have seen, the dispositionalist is not in fact threatened by the affirmative reply, even

if he accepts it as correct: provided that the subject's linguistic ability *was* exercised before and after the injury, the dispositionalist is under no pressure to deny that it existed, but could not be exercised, *during* the injury. It seems, however, that the dispositionalist can now create a problem for the position Chomsky himself represents—a problem, that is, for the position that linguistic knowledge *can* be characterised at all without *any* reference to its behavioural manifestations.

The dispositionalist would first observe that if, as Chomsky believes, linguistic knowledge could be characterised independently of any reference to its behavioural manifestations, then, in the description of Chomsky's thought experiment, the reference to the subject's pre-injury and post-recovery behavioural manifestations of its knowledge of English should be *inessential*: everything we would say about the subject's knowledge of English "during the intervening period" we should also be able to say *without* taking into account the fact that it did behaviourally manifest that knowledge before the injury and did behaviourally manifest the same knowledge after the recovery. But this radical separation of attributions of knowledge from references to behavioural manifestations—the dispositionalist would object—does not appear to be in fact possible, as we can appreciate by considering a different thought experiment, which abstracts away from precisely those factors that ought to be inessential on Chomsky's account:

> A child is born to a couple of English speaking parents. A few weeks after its birth—and so, before the child is in a position to use English or any other natural language—it suffers cerebral damage which, though in all other respects does not prevent the natural growth of his brain, completely prevents, for ten consecutive years, the use of English or of any other natural language "in speech, comprehension, or let us suppose even in thought". On its tenth birthday—and so, while the effects of the injury are still fully in place—the child suddenly dies. Did the child know English or any other natural language at any point in its short life?

The answer to *this* question is, obviously, negative. A child who, at no point in its entire life, *used* a language—has never spoken it, has never understood it, has never even thought in it—cannot be said to *know* that language, no matter what the architecture of its brain is: even if a post-mortem examination of the child in question was to reveal that the so-called "language centres" in its brain (that is, the brain structures that are supposed to be linguistically relevant but strictly distinct from whatever brain structures are dedicated to the management of language *use*) were as well developed as anyone else's in its community, the fact that the reputed "centres" were, in view of the child's total incapacity for linguistic

behaviour, *never* in its life connected to the production or understanding of any element of English or of any other natural language would surely suffice for concluding that the child never knew English or some other natural language. But this obviously true negative answer—the dispositionalist would observe—is clearly not one that Chomsky's position would allow him to acknowledge. For the whole point of Chomsky's anti-dispositionalist campaign is precisely to assert that a person's *knowledge* of a language simply consists in the existence of certain specialised structures in its brain that are distinct from whatever brain structures may be dedicated to the management of language use, and can be fully characterised without reference to any aspect of the person's behaviour (for example, producing sentences, understanding sentences, etc.) that constitutes language use. Assuming, therefore, that Chomsky would not want to concede that direct or indirect reference to language use is necessary for ascriptions of linguistic knowledge (a concession that would make his position, in all relevant respects, indistinguishable from the dispositionalist's), the only option available to him when confronted with the above thought experiment would be to claim that, if the "language centres" in this child's brain were found, after a post-mortem examination, to be as well developed as anyone else's in its community, the child should indeed be credited with having possessed a great amount of linguistic *knowledge*, even though it was, throughout its entire life, fully unable to either produce or understand anything occurring in any natural language whatsoever. But this only shows—the dispositionalist would conclude—that if Chomsky was to maintain his position, he would have to use a concept of "knowledge" that is so idiosyncratic and obscure that any attempt at further communication with him on these matters would be bound to fail (as, in fact, many philosophers have, on independent grounds, long ago suspected that it would; cf., among others, Nagel 1969, Stich 1971, Cooper 1975).

It seems to me that this argument is very difficult to counter, and certainly impossible to ignore. If so, the conclusion we are finally entitled to draw is stronger than the one previously derived: not only has Chomsky failed to produce a valid argument against the dispositional construal of linguistic knowledge, but it is also the case that the dispositionalist can produce a very effective argument of his own against Chomsky's anti-dispositionalist construal. To put it concisely, the relative strengths of the dispositionalist and the anti-dispositionalist conceptions of linguistic knowledge appear to be exactly the opposite of what Chomsky thought they were.

References

Alston, William P. 1964. Linguistic acts. *American Philosophical Quarterly* 1. 138–146.

Alston, William P. 1994. Illocutionary acts and linguistic meaning. In Savas L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, 29–49. London & New York: Routledge.

Alston, William. P. 2000. *Illocutionary acts and sentence meaning*. Ithaca, NY: Cornell University Press.

Alston, William P. 2007. Illocutionary acts and truth. In Dirk Greimann & Geo Siegwart (eds.), *Truth and speech acts: Studies in the philosophy of language*, 9–30. London & New York: Routledge.

Apel, Karl-Otto. 1991. Is intentionality more basic than linguistic meaning? In Ernest Lepore & Robert Van Gulick (eds.), *John Searle and his critics*, 31–55. Oxford: Blackwell.

Apel, Karl-Otto. 1994. *Selected essays, Vol.1: Towards a transcendental semiotics*. Atlantic Highlands, NJ: Humanities Press.

Atlas, Jay David. 1989. *Philosophy without ambiguity: A logico-linguistic essay*. Oxford: Clarendon Press.

Austin, J. L. 1946. Other minds. *Proceedings of the Aristotelian Society*, Supplementary Volume 20. 148–187.

Austin, J. L. 1950. Truth. *Proceedings of the Aristotelian Society*, Supplementary Volume 24. 111–128.

Austin, J. L. 1963. Performative-Constative. In Charles E. Caton (ed.), *Philosophy and ordinary language,* 22–33. Chicago: University of Illinois Press. (Originally published as "Performatif-Constatif" in *Cahiers de Royaumont*, Philosophie No. IV, *La philosophie analytique*, Paris: Minuit, 1962.)

Austin J. L. 1975. *How to do things with words*, 2nd edn., ed. by J. O. Urmson & Marina Sbisà. Oxford: Oxford University Press. (1st edn., 1962.)

Austin J. L. 1979. *Philosophical papers*, 3rd edn., ed. by J. O. Urmson & G. J. Warnock. Oxford: Oxford University Press. (1st edn., 1961.)

Bach, Kent. 1975. Performatives are statements too. *Philosophical Studies* 28. 229–236.

Bach, Kent & Robert M. Harnish. 1979. *Linguistic communication and speech acts*. Cambridge, Mass.: The MIT Press.

Barnes, Jonathan (ed.) 1984. *The complete works of Aristotle: The revised Oxford translation*, Vols. 1 and 2. Princeton, NJ: Princeton University Press.

Barsalou, Lawrence W. 1983. *Ad hoc* categories. *Memory and Cognition* 11. 211–227.

Barsalou, Lawrence W. 1987. The instability of graded structure: Implications for the nature of concepts. In Ulrich Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorisation*, 101–140. Cambridge: Cambridge University Press.

Barsalou, Lawrence W. & Douglas L. Medin. 1986. Concepts: Fixed definitions or context-dependent representations? *Cahiers de Psychologie Cognitive* 6. 187–204.

Bendix, Edward H. 1966. *Componential analysis of general vocabulary: The semantic structure of a set of verbs in English, Hindi, and Japanese*. Bloomington: Indiana University Press.

Bennett, Jonathan. 1991. How do gestures succeed? In Ernest Lepore and Robert van Gulick (eds.), *John Searle and his critics*, 3–15. Oxford: Blackwell.

Binkley, Timothy. 1979. The principle of expressibility. *Philosophy and Phenomenological Research* 39. 307–325.

Bird, Graham H. 1994. Relevance theory and speech acts. In Savas L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, 292–311. London & New York: Routledge.

Blome-Tillmann, Michael. 2008. Conversational implicature and the cancellability test. *Analysis* 68. 156–160.

Borge, Steffen. 2009. Conversational implicatures and cancellability. *Acta Analytica* 24. 149–154.

Brandom, Robert B. 1983. Asserting. *Noûs* 17. 637–650.

Brandom, Robert B. 1994. *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, Mass.: Harvard University Press.

Brandom, Robert B. 2000. *Articulating reasons: An introduction to inferentialism*. Cambridge, Mass.: Harvard University Press.

Brandom, Robert B. 2008. *Between saying and doing: Towards an analytic pragmatism.* Oxford: Oxford University Press.

Brandom, Robert B. 2009. *Reason in philosophy: Animating ideas*. Cambridge, Mass.: Harvard University Press.

Burge, Tyler. 1979. Individualism and the mental. *Midwest Studies in Philosophy* 4. 73–121.

Carston, Robyn. 2002. *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.

Carston, Robyn. 2010. Explicit communication and 'free' pragmatic enrichment. In Belén Soria & Esther Romero (eds.), *Explicit communication: Robyn Carston's pragmatics*, 217–285. New York: Palgrave Macmillan.

Cappelen, Herman & Ernie Lepore. 1997a. Varieties of quotation. *Mind* 106. 429–450.

Cappelen, Herman & Ernie Lepore. 1997b. On an alleged connection between indirect speech and the theory of meaning. *Mind and Language* 12. 278–296.

Cappelen, Herman & Ernie Lepore 1998. Reply to Tsohatzidis. *Mind* 107. 665–666.

Chalmers, David J. 1996. *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/semantics interface. In Adriana Belletti (ed.), *Structures and beyond*, 39–103. Oxford: Oxford University Press.

Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press.

Chierchia, Gennaro. 2017. Scalar implicatures and their interface with grammar. *Annual Review of Linguistics* 3. 245–264.

Chierchia, Gennaro, Danny Fox & Benjamin Spector. 2012. Scalar implicature as a grammatical phenomenon. In Claudia Maierborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of contemporary research, Vol. III*, 2297–2331. Berlin & Boston: De Gruyter.

Chomsky, Noam. 1980. *Rules and representations*. New York, NY: Columbia University Press.

Chomsky, Noam. 1988. *Language and problems of knowledge: The Managua lectures*. Cambridge, Mass.: The MIT Press.

Chomsky, Noam. 1992. Language and interpretation: philosophical reflections and empirical inquiry. In John Earman (ed.), *Inference, explanation, and other frustrations: Essays in the philosophy of science*, 99–128. Berkeley, CA: University of California Press.

Cohen, L. Jonathan. 1979. The semantics of metaphor. In Andrew Ortony (ed.), *Metaphor and thought*, 64–77. Cambridge: Cambridge University Press.

Cohen, L. Jonathan & Avishai Margalit. 1972. The role of inductive reasoning in the interpretation of metaphor. In Donald Davidson & Gilbert Harman (eds.), *Semantics of natural language*, 722–740. Dordrecht: Reidel.

Cooper, David E. 1975. *Knowledge of language*. London: Prism Press.

Cooper, David E. 1986. *Metaphor*. Oxford: Blackwell.

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. Alan. 1992. Reply to Savas L. Tsohatzidis, "A fake typicality Constraint on asymmetric acceptability". *Cognitive Linguistics* 3-4: 399–401.

Culicover, Peter W. 1992. English tag questions in Universal Grammar. *Lingua* 88. 193–226.

Davidson, Donald. 1967. Truth and meaning. *Synthese* 17. 304–323.

Davidson, Donald. 1968. On saying that. *Synthese* 19. 130–146.

Davidson, Donald. 1973. Radical interpretation. *Dialectica* 27. 313–328.

Davidson, Donald. 1976. Reply to Foster. In Gareth Evans & John McDowell (eds.), *Truth and meaning: Essays in semantics*, 33–41. Oxford: Clarendon Press.

Davidson, Donald. 1978. What metaphors mean. *Critical Inquiry* 5. 31–47.

Davidson, Donald. 1979. Quotation. *Theory and Decision* 11. 27–40.

Davidson, Donald. 1980. *Essays on actions and events*. Oxford: Oxford University Press.

Davidson, Donald. 1984. *Inquiries into truth and interpretation*. Oxford: Oxford University Press.

Davidson, Donald. 1999. Reply to Cappelen and Lepore. In Ursula M. Zeglen (ed.), *Donald Davidson: Truth, meaning and knowledge*, 100–101. London: Routledge.

Davidson, Donald. 2005. *Truth and predication*. Cambridge, Mass.: Harvard University Press.

Davis, Steven. 1979. Perlocutions. *Linguistics and Philosophy* 3. 225–243.

Davis, Wayne A. 1998. *Implicature: Intention, convention and principle in the failure of Gricean theory*. Cambridge: Cambridge University Press.

Dennett, Daniel C. 1990. The myth of original intentionality. In K. A. Mohyeldin Said, W. H. Newton-Smith, R. Viale & K. V. Wilkes (eds.), *Modelling the mind*, 43–62. Oxford: Oxford University Press.

Devitt, Michael. 1990. Meanings just ain't in the head. In George Boolos (ed.), *Meaning and method: Essays in honor of Hilary Putnam*, 79–104. Cambridge: Cambridge University Press.

Devlin, Keith. 1991. *Logic and information*. Cambridge: Cambridge University Press.

Dreyfus, Hubert L. 2001. Phenomenological description versus rational reconstruction. *Revue Internationale de Philosophie* 55. 181–196.

Dummett, Michael. 1973. *Frege: Philosophy of language*. London: Duckworth.

Dummett, Michael. 1991. *The logical basis of metaphysics*. Cambridge, Mass.: Harvard University Press.

Dummett, Michael. 2004. *Truth and the past*. New York: Columbia University Press.

Dummett, Michael. 2007. Reply to Hilary Putnam. In Randall E. Auxier & Lewis Edwin Hahn (eds.), *The philosophy of Michael Dummett*, 168–184. Chicago: Open Court.

Dummett, Michael. 2010. *The nature and future of philosophy*. New York: Columbia University Press.

Faigenbaum, Gustavo. 2001. *Conversations with John Searle*. Montevideo: Libros en Red.

Fine, Kit. 2012a. The pure logic of ground. *The Review of Symbolic Logic* 5. 1—25.

Fine, Kit. 2012b. Guide to ground. In Fabrice Correia & Benjamin Schnieder (eds.), *Metaphysical grounding: Understanding the structure of reality*, 37–80. Cambridge: Cambridge University Press.

Fitzpatrick, Dan. 2003. Searle and collective intentionality: The self-defeating nature of internalism with respect to social facts. *American Journal of Economics and Sociology* 62. 45–66.

Fogelin, Robert J. 1989. *Figuratively speaking*. New Haven, CT: Yale University Press.

Forguson, L. W. 1973. Locutionary and illocutionary acts. In Isaiah Berlin (ed.) *Essays on J. L. Austin*, 160–185. Oxford: Clarendon Press.

Frege, Gottlob. 1918. Der Gedanke, eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus* 1. 58–77.

Frye, Marilyn. 1976. On saying. *American Philosophical Quarterly* 13. 123–127.

Furberg, Mats. 1971. *Saying and meaning: A main theme in J. L. Austin's philosophy*. Oxford: Blackwell.

Gazdar, Gerald. 1981. Speech act assignment. In Aravind K. Joshi, Bonnie L. Webber & Ivan A. Sag (eds.), *Elements of discourse understanding*, 64–83. Cambridge: Cambridge University Press.

Grewendorf, Günther & Georg Meggle (eds.). 2002. *Speech acts, mind and social reality: Discussions with John R. Searle*. Dordrecht: Kluwer Academic Publishers.

Grice, Paul. 1957. Meaning. *The Philosophical Review* 66. 377–388.

Grice, Paul. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics, Vol. 3: Speech acts*, 41–58. New York: Academic Press.

Grice, Paul. 1989. *Studies in the way of words*. Cambridge, Mass.: Harvard University Press.

Griffiths, Leslie. 1969. The logic of Austin's locutionary subdivision. *Theoria* 35. 204–214.

Hamblin, C. L. 1973. Questions in Montague English. *Foundations of Language* 10. 41–53.

Hazlett, Allan. 2010. The myth of factive verbs. *Philosophy and Phenomenological Research* 80. 497–522.

Haegeman, Liliane. 2003. Conditional clauses: external and internal syntax. *Mind and Language* 18. 317–339.

Han, Chung-Hye. 2000. *The structure and interpretation of imperatives: Mood and force in universal grammar*. New York: Garland.

Harnish, Robert M. 1990. Speech acts and intentionality. In Armin Burkhardt (ed.), *Speech acts, meaning and intentions: Critical approaches to the philosophy of John R. Searle*, 169–193. Berlin & New York: De Gruyter.

Harnish, Robert M. 1994. Mood, meaning and speech acts. In Savas L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, 407–459. London & New York: Routledge.

Higginbotham, James. 1996. The semantics of questions. In Shalom Lappin (ed.), *The handbook of contemporary semantic theory*, 361–84. Oxford: Blackwell.

Holdcroft, David. 1976. Forms of indirect communication: an outline. *Philosophy and Rhetoric* 9. 147–161.

Holdcroft, David. 1978. *Words and deeds: Problems in the theory of speech acts*. Oxford: Clarendon Press.

Holton, Richard. 1997. Some telling examples: A reply to Tsohatzidis. *Journal of Pragmatics* 28. 625–628.

Hornsby, Jennifer. 2006. Speech acts and performatives. In Ernie Lepore & Barry C. Smith (eds.), *The Oxford handbook of philosophy of language*, 893–909. Oxford: Oxford University Press.

Humberstone, I. L. 1992. Direction of fit. *Mind* 101. 59–83.

Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, Mass.: The MIT Press.

Jary, Mark. 2010. *Assertion*. New York: Palgrave Macmillan.

Kaplan, David. 1989a. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 481–564. Oxford: Oxford University Press.

Kaplan, David. 1989b. Afterthoughts. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 565–614. Oxford: Oxford University Press.

Karttunen, Lauri. 1978. Syntax and semantics of questions. In Henry Hiz (ed.), *Questions*, 165–210. Dordrecht: Reidel.

Kasher, Asa. 1976. Conversational maxims and rationality. In Asa Kasher (ed.), *Language in focus: Foundations, methods, and systems*, 197–216. Dordrecht: Reidel.

Kenny, Anthony. 1975. *Will, freedom, and power*. Oxford: Blackwell.

Kenny, Anthony. 1984. *The legacy of Wittgenstein*. Oxford: Blackwell.

Kenny, Anthony. 1989. *The metaphysics of mind*. Oxford: Oxford University Press.

Kittay, Eva Fedder. 1987. *Metaphor: Its cognitive force and linguistic structure*. Oxford: Oxford University Press.

Koepsell, David & Lawrence S. Moss (eds.). 2003. *John Searle's ideas about social reality: Extensions, criticisms, and reconstructions*. Oxford: Blackwell.

Krifka, Manfred. 2001. Quantifying into question acts. *Natural Language Semantics* 9. 1–40.

Kripke, Saul A. 1972. Naming and necessity. In Gilbert Harman & Donald Davidson (eds.), *Semantics of natural language*, 253–355, 763–9. Dordrecht: Reidel.

Kripke, Saul A. 1980. *Naming and necessity*. Cambridge, Mass.: Harvard University Press.

Kvanvig, Jonathan L. 2006. *The knowability paradox*. Oxford: Oxford University Press.

Lakoff, George. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lehrer, Adrienne. 1990. Prototype theory and its implications for lexical analysis. In Savas L. Tsohatzidis (ed.), *Meanings and prototypes: Studies in linguistic categorisation*, 368–381. London & New York: Routledge.

Lepore, Ernest & Brian McLaughlin (eds.). 1985. *Actions and events: Perspectives on the philosophy of Donald Davidson*. Oxford: Blackwell.

Lepore, Ernest & Robert van Gulick (eds.). 1991. *John Searle and his critics*. Oxford: Blackwell.

Levine, Joseph. 2001. *Purple haze: The puzzle of consciousness*. Oxford: Clarendon Press.

Lewis, David K. 1969. *Convention*. Cambridge, Mass.: Harvard University Press.

Lewis, David K. 1975. Language and languages. In Keith Gunderson (ed.), *Language, mind and knowledge*, 3–35. Minneapolis, MN: University of Minnesota Press.

Lewis, David K. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8. 339–359.

Liedtke, Frank W. 1990. Representational semantics and illocutionary acts. In Armin Burkhardt (ed.), *Speech acts, meaning and intentions: Critical approaches to the philosophy of John R. Searle*, 194–209. Berlin & New York: De Gruyter.

MacCormack, Earl R. 1985. *A cognitive theory of metaphor*. Cambridge, MA: MIT Press.

MacFarlane, John. 2011. What is assertion? In Jessica Brown & Herman Cappelen (eds.), *Assertion: New philosophical essays*, 79–96. Oxford: Oxford University Press.

Malcolm, Norman. 1991. 'I believe that p'. In Ernest Lepore & Robert Van Gulick (eds.), *John Searle and his critics*, 159–168. Oxford: Blackwell.

McCulloch, Gregory. 1995. *The mind and its world*. London & New York: Routledge.

Medin, Douglas L. & Lawrence W. Barsalou. 1987. Categorisation processes and categorical perception. In Steven Harnad (ed.), *Categorical perception: The groundwork of cognition*, 455–490. Cambridge: Cambridge University Press.

Meggle, Georg (ed.). 2002. *Social facts and collective intentionality*. Frankfurt: Dr. Hansel-Hohenhausen AG.

Millikan, Ruth Garrett. 2005. *Language: A biological model*. Oxford: Oxford University Press.

Mills, Stephen. 1993. Wittgenstein and connectionism: A significant complementarity? In Christopher Hookway & Donald M. Peterson (eds.), *Philosophy and cognitive science*, 137–157.Cambridge: Cambridge University Press.

Mulligan, Kevin (ed.). 1987. *Speech act and Sachverhalt: Reinach and the foundations of realist phenomenology*. Dordrecht: Martinus Nijhoff.

Munsat, Stanley. 1986. Wh-complementisers. *Linguistics and Philosophy* 9. 191–216.

Nagel, Thomas. 1969. Linguistics and epistemology. In Sidney Hook (ed.), *Linguistics and philosophy*, 171–182. New York: New York University Press.

Pagin, Peter. 2004. Is assertion social? *Journal of Pragmatics* 36. 833–59.

Pagin, Peter. 2008. Assertion. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy*. Fall 2008 Edition. Stanford, CA: Center for the Study of Language and Information.

Pagin, Peter. 2011. Information and assertoric force. In Jessica Brown & Herman Cappelen (eds), *Assertion: New philosophical essays*, 97–135. Oxford: Oxford University Press.

Preston, John and Mark Bishop (eds.). 2002. *Views into the Chinese room: New Essays on Searle and artificial intelligence*. Oxford: Clarendon Press.

Prichard, H. A. 2002. *Moral writings*, ed. by J. MacAdam. Oxford: Oxford University Press.

Progovac, Ljiljana. 1999. Events and economy of coordination. *Syntax* 2. 141–159.

Putnam, Hilary. 1975. The meaning of 'meaning'. In Keith Gunderson (ed.), *Language, mind and knowledge*, 131–193. Minneapolis, MN: University of Minnesota Press.

Quine, W.V. O. 1970. Methodological reflections on current linguistic theory. *Synthese* 21. 386–398.

Récanati, François. 1980. Qu'est-ce qu'un acte locutionnaire? *Communications* 32. 190–215.

Récanati, François. 1981. *Les enoncés performatifs: Contribution à la pragmatique*. Paris: Minuit.

Récanati, François. 1987. *Meaning and force: The pragmatics of performative utterances*. Cambridge: Cambridge University Press.

Récanati, François. 2003. The limits of expressibility. In Barry Smith (ed.), *John Searle*, 189–213. Cambridge: Cambridge University Press.

Reimer, Marga. 1998. What is meant by 'what is said': a reply to Cappelen and Lepore. *Mind and Language* 13. 598–604.

Richard, Mark. 1998. Semantic theory and indirect speech. *Mind and Language* 13. 605–616.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of grammar: Handbook of generative syntax*, 281–337. Dordrecht: Kluwer.

Robinson, Jenefer. 1974. The individuation of speech acts. *The Philosophical Quarterly* 24. 316–336.

Ross, J. F. 1981. *Portraying analogy*. Cambridge: Cambridge University Press.

Ruben, David-Hillel. 1997. John Searle's *The Construction of Social Reality*. *Philosophy and Phenomenological Research* 57: 443–447.

Sadock, Jerrold M. 1994. Toward a grammatically realistic typology of speech acts. In Savas L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, 393–406. London & New York: Routledge.

Sampson, Geoffrey. 1982. The economics of conversation. In Neil V. Smith (ed.), *Mutual knowledge*, 200–210. New York: Academic Press.

Schiffer, Stephen R. 1972. *Meaning*. Oxford: Clarendon Press.

Schnieder, Benjamin. 2011. A logic for 'because'. *The Review of Symbolic Logic* 4. 445—465.

Schuhmann, Karl & Barry Smith. 1990. Elements of speech act theory in the philosophy of Thomas Reid. *History of Philosophy Quarterly* 7. 47–66.

Searle, John R. 1958. Proper names. *Mind* 67. 166–173.

Searle, John R. 1964. How to derive 'ought' from 'is'. *The Philosophical Review* 73. 43–58.

Searle, John R. 1968. Austin on locutionary and illocutionary acts. *The Philosophical Review* 77. 405–424.

Searle, John R. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, John R. 1975a. A taxonomy of illocutionary acts. In Keith Gunderson (ed.), *Language, mind and knowledge*, 344–369. Minneapolis, MN: University of Minnesota Press.

Searle, John R. 1975b. Indirect speech acts. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics, Vol. 3: Speech acts*, 59–82. New York: Academic Press.

Searle, John R. 1978. *Prima facie* obligations. In Joseph Raz (ed.), *Practical reasoning*, 81–90. Oxford: Oxford University Press.

Searle, John R. 1979. *Expression and meaning: Studies in the theory of speech acts*. Cambridge: Cambridge University Press.

Searle, John R. 1980. Minds, brains, and programs. *The Behavioral and Brain Sciences* 3. 417–424.

Searle, John R. 1983. *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.

Searle, John R. 1984. *Minds, brains and science: The 1984 Reith lectures*. Cambridge, Mass.: Harvard University Press.

Searle, John R. 1986a. Meaning, communication, and representation. In Richard Grandy & Richard Warner (eds.), *Philosophical grounds of rationality: Intentions, categories, ends*, 209–226. Oxford: Clarendon Press.

Searle, John R. 1986b. Notes on conversation. In Donald G. Ellis & William A. Donohue (eds.), *Contemporary issues in language and discourse processes*, 7–19. Hillsdale, NJ: Lawrence Erlbaum Associates.

Searle, John R. 1989. How performatives work. *Linguistics and Philosophy* 12. 535–558.

Searle, John R. 1990. Consciousness, explanatory inversion, and cognitive science. *The Behavioral and Brain Sciences* 13. 585–596.

Searle, John R. 1991a. Response: meaning, intentionality, and speech acts. In Ernest Lepore & Robert Van Gulick (eds.), *John Searle and his critics*, 81–102. Oxford: Blackwell.

Searle, John R. 1991b. Response: perception and the satisfactions of intentionality. In Ernest Lepore & Robert Van Gulick (eds.), *John Searle and his critics*, 181–192. Oxford: Blackwell.

Searle, John R. 1991c. Intentionalistic explanations in the social sciences. *Philosophy of the Social Sciences* 21. 332–344.

Searle, John R. 1992. *The rediscovery of the mind*. Cambridge, Mass.: The MIT Press.

Searle, John R. 1995. *The construction of social reality*. New York: Free Press.

Searle, John R. 1997a. *The mystery of consciousness*. London: Granta Books.

Searle, John R. 1997b. Responses to critics of *The Construction of Social Reality*. *Philosophy and Phenomenological Research* 57. 449–458.

Searle, John R. 1998. *Mind, language and society: Philosophy in the real world*. New York: Basic Books.

Searle, John R. 1999. Fact and value, 'is' and 'ought', and reasons for action. In G. O. Mazur (ed.), *Twenty-five year commemoration to the life of Hans Kelsen (1898–1973)*, 20–34. New York: Semenenko Foundation.

Searle, John R. 2001. *Rationality in action*. Cambridge, Mass.: The MIT Press.

Searle, John R. 2002. *Consciousness and language*. Cambridge: Cambridge University Press.

Searle, John R. 2004. *Mind: A brief introduction*. Oxford: Oxford University Press.

Searle, John R. 2007. *Freedom and neurobiology: Reflections on free will, language, and political power*. New York: Columbia University Press.

Searle, John R. 2010. *Making the social world: The structure of human civilisation*. Oxford: Oxford University Press.

Searle, John R. & Daniel Vanderveken. 1985. *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Sennet, Adam. 2013. Parataxis. In Ernie Lepore and Kirk Ludwig (eds.), *A companion to Donald Davidson*, 191–207. Oxford: Wiley Blackwell.

Siebel, Mark. 2001. Searle's representing account of illocutionary acts and its weak spots. *Acta Philosophica Fennica* 69. 97–111.

Smith, Barry. 1990. Toward a history of speech act theory. In Armin Burkhardt (ed.), *Speech acts, meaning and intentions: Critical approaches to the philosophy of John Searle*, 29–61. Berlin & New York: De Gruyter.

Smith, Barry (ed.). 2003. *John Searle*. Cambridge: Cambridge University Press.

Soames, Scott. 2003. *Philosophical analysis in the twentieth century, Vol. 2: The age of meaning*. Princeton, NJ: Princeton University Press.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Blackwell.

Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and cognition*. Second Edition. Oxford: Blackwell.

Stern, Josef. 1985. Metaphor as demonstrative. *Journal of Philosophy* 82. 677–710.

Stern, Josef. 2000. *Metaphor in context*. Cambridge, Mass.: The MIT Press.

Stern, Josef. 2006. Figurative language. In Michael Devitt & Richard Hanley (eds.), *The Blackwell guide to the philosophy of language*, 168–185. Oxford: Blackwell.

Stern, Josef. 2011. Metaphor and minimalism. *Philosophical Studies* 153. 273–298.

Stjernberg, Fredrik. 2009. Restricting factiveness. *Philosophical Studies* 146. 29–48.

Stich, Stephen P. 1971. What every speaker knows. *The Philosophical Review* 80. 476–496.

Strawson, P. F. 1964. Intention and convention in speech acts. *The Philosophical Review* 73. 439–460.

Strawson, P. F. 1973. Austin and 'locutionary meaning'. In Isaiah Berlin (ed.), *Essays on J. L. Austin*, 46–68. Oxford: Clarendon Press.

Stroll, Avrum. 1989. What water is, or back to Thales. *Midwest Studies in Philosophy* 14. 258–274.

Stroll, Avrum. 1998. Proper names, names, and fictive objects. *The Journal of Philosophy* 95. 522–534.

Stroll, Avrum. 2000. *Twentieth-century analytic philosophy*. New York: Columbia University Press.

Tarski, Alfred. 1944. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4. 341–376.

Taylor, Talbot J. & Deborah Cameron. 1987. *Analysing conversation: Rules and units in the structure of talk*. Oxford: Pergamon Press.

Thagard, Paul. 2000. *Coherence in thought and action*. Cambridge, Mass.: The MIT Press.

Thalos, Mariam. 2003. Searle's Foole: How a constructionist account of society cannot substitute for a causal one. *American Journal of Economics and Sociology* 62. 105–122.

Thau, Stewart. 1972. The distinction between rhetic and illocutionary acts. *Analysis* 32. 177–183.

Tsohatzidis, Savas L. 1992a. A fake typicality constraint on asymmetric acceptability. *Cognitive Linguistics* 3-4: 391–398

Tsohatzidis, Savas L. 1992b. Fake means fake: A reply to D. A. Cruse's reply *Cognitive Linguistics* 3-4: 403–408.

Tsohatzidis, Savas L. 1993a. Speaking of truth-telling: The view from *wh*-complements. *Journal of Pragmatics* 19. 271–279.

Tsohatzidis, Savas L. 1993b. Emotional states and linguistic events: a study of conceptual misconnections. *Pragmatics and Cognition* 1: 229–242.

Tsohatzidis, Savas L. (ed.). 1994. *Foundations of speech act theory: Philosophical and linguistic perspectives*. London & New York: Routledge.

Tsohatzidis, Savas L. 1997. More telling examples: A response to Holton. *Journal of Pragmatics* 28. 629–636.

Tsohatzidis, Savas L. 1998. The hybrid theory of mixed quotation. *Mind* 107. 661–664.

Tsohatzidis, Savas L. 2003. Lost hopes and mixed quotes. *Belgian Journal of Linguistics* 17. 213–229.

Tuomela, Raimo. 2002. *Philosophy of social practices: A collective acceptance view*. Cambridge: Cambridge University Press.

Turner, Stephen P. 1999. Searle's social reality. *History and Theory* 38. 211–231.

Urmson, J. O., W. V. O. Quine & Stuart Hampshire. (1969). A symposium on Austin's method. In K. T. Fann (ed.), *Symposium on J. L. Austin*, 76–97. London & New York: Routledge.

Vanderveken, Daniel. 1991. Non-literal speech acts and conversational maxims. In Ernest Lepore & Robert Van Gulick (eds.), *John Searle and his critics*, 371–384. Oxford: Blackwell.

Vanderveken, Daniel & Susumu Kubo (eds.) 2002. *Essays in speech act theory*. Amsterdam: John Benjamins.

Vendler, Zeno. 1972. *Res cogitans*. Ithaca, NY: Cornell University Press.

Vendler, Zeno. 1978. Escaping from the cave. *Canadian Journal of Philosophy* 8. 79–87.

Vendler, Zeno. 1980. Telling the facts. In John R. Searle, Ferenc Kiefer & Manfred Bierwisch (eds.)*, Speech act theory and pragmatics*, 273–290. Dordrecht: Reidel.

Vermazen, Bruce & Merill B. Hintikka (eds.) 1985. *Essays on Davidson: Actions and events*. Oxford: Oxford University Press.

Viskovatoff, Alex. 2002. Searle's Background. *Journal of Economic Methodology* 9. 65–80.

Viskovatoff, Alex. 2003. Searle, rationality and social reality. *American Journal of Economics and Sociology* 62. 7–44.

Warnock, G. J. 1983. *Morality and language*. Oxford: Blackwell.

Weiner, Matthew. 2006. Are all conversational implicatures cancellable? *Analysis* 66. 127—30.

Weinreich, Uriel. 1963. On the semantic structure of language. In Greenberg, G.H. (ed.), *Universals of language*. Cambridge, MA: The MIT Press.

Weinreich, Uriel. 1972. *Explorations in semantic theory*. The Hague: Mouton.

Welbourne, Michael. 1989. A puzzle about telling. *Philosophy* 64. 175–185.

Wettersten, John. 1998. The analytic study of social ontology: breakthrough or cul-de-sac? *Philosophy of the Social Sciences* 28. 132–151.

Wilson, Deirdre & Dan Sperber. 2012. *Meaning and relevance*. Cambridge: Cambridge University Press.

Wilson, Jessica M. 2014. No work for a theory of grounding. *Inquiry* 57. 535–579.

Williamson, Timothy. 1996. Knowing and asserting. *The Philosophical Review* 105. 489–523.

Williamson, Timothy. 2000. *Knowledge and its limits*. Oxford: Oxford University Press.

Williamson, Timothy. 2009. Replies to critics. In Patrick Greenough & Duncan Pritchard (eds.), *Williamson on knowledge*, 279–384. Oxford: Oxford University Press.

Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Blackwell.

Wittgenstein, Ludwig. 1980. *Remarks on the philosophy of psychology,* Vols. 1 and 2. Oxford: Blackwell.

Yablo, Stephen. 1993. Paradox without self-reference. *Analysis* 53. 251–252.

Index

https://doi.org/10.1515/9783110687538-27