

5-2015

Representation without Thought: Confusion, Reference, and Communication

Elmar Geir Unnsteinsson

Graduate Center, City University of New York

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

 Part of the [Philosophy Commons](#)

Recommended Citation

Unnsteinsson, Elmar Geir, "Representation without Thought: Confusion, Reference, and Communication" (2015). *CUNY Academic Works*.

https://academicworks.cuny.edu/gc_etds/1168

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.

REPRESENTATION WITHOUT THOUGHT: CONFUSION, REFERENCE, AND
COMMUNICATION

by

ELMAR GEIR UNNSTEINSSON

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the
requirements for the degree of Doctor of Philosophy, The City University of New York

2015

© 2015

ELMAR GEIR UNNSTEINSSON

All rights reserved

This manuscript has been read and accepted for the
Graduate Faculty in Philosophy in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Michael Devitt

Date

Chair of Examining Committee

John Greenwood

Date

Executive Officer

Michael Devitt

Gary Ostertag

David Rosenthal

David Papineau

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

REPRESENTATION WITHOUT THOUGHT: CONFUSION, REFERENCE, AND COMMUNICATION

by

Elmar Geir Unnsteinsson

Adviser: Professor Stephen Neale

Philosophy of mind has been dominated, since Frege, by a puzzle-driven methodology. This tradition aims to provide a coherent system for describing specific semantic features of all conceivable cases where the speaker is confused about the identity of an object. The first chapter develops a theory on which confused identity is a mental state of an agent who either believes falsely *that* $a = b$ or believes falsely *that* $a \neq b$. Many influential arguments in philosophy are puzzle-driven; Kripke on semantic and speaker reference, Reimer and Kaplan on demonstratives. I show in detail how these and other arguments are invalidated because of doubtful assumptions about confused identity.

The alternative is ‘explanation-driven semantics.’ Combining Gricean intentionalism and teleose- mantic ideas—which are usually thought to be in strict opposition—I show that the basic task of a theory of meaning is to explain how humans express and communicate their thoughts so successfully by linguistic means. Puzzle-driven semantics has no relevance to this project. Confused speakers are ‘abnormal’ in Millikan’s sense: their mental state disrupts the proper function of the relevant singular terms in their idiolect or language of thought.

My positive theory defines a notion of ‘edenic reference,’ which idealizes away from confusion in defining the proper function of singular terms. Speakers must satisfy certain cog-

tive constraints if their utterances are to have a role in explaining the maintenance of a practice of using a singular term in a population. A related constraint on coreference states, roughly, that if a speaker utters a simple sentence containing more than one singular term, she cannot be indifferent as to whether they are intended to refer to the same thing or to distinct things. Such indifference is not impossible, but it disturbs the proper function of the linguistic construction—or what Grice called the ‘optimal’ mental state with respect to a form of linguistic behavior.

When I first read it, I thought it was an exercise in irony. Then a very skilful parody of a certain attitude. Then I realised it was serious—it was at the moment I searched my memory and rooted out certain fantasies of my own. But what seemed to me important was that it could be read as parody, irony or seriously. It seems to me this fact is another expression of the fragmentation of everything, . . . linked with what I feel to be true about language, the thinning of language against the density of our experience.

DORIS LESSING

Preface

There appears to be a never-ending list of questions and issues about language, meaning, and representation which are both intrinsically worthy of investigation and, at present, seemingly intractable, not to say mystifying. I do not, however, endorse Chomsky's well-known edict that one ought to put mysteries to rest and focus only on allegedly solvable problems. Such a position tends to obscure the larger, more foundational concerns that should be driving our inquiry in the first place. Still, I do not wish to jettison the distinction between so-called mysteries and problems, only to propose that tackling the latter always requires acknowledgment of the former.

The foundational questions, which got me interested in philosophy of language to begin with, include such luminaries as, What is linguistic meaning? How are semantic facts grounded in other more basic facts? How is it that, by and large, we can communicate highly complicated contents so effortlessly and efficiently by the use of language? How and why did natural languages evolve? How are they learned?

One would think that the philosopher is charged with keeping these questions alive and relevant, if anyone is. A major lesson of this dissertation, however, is that even the philosophers have lost touch with foundational concerns in the study of language. I argue that philosophers of language have pursued semantic puzzles—specifically ones involving speakers who are confused about the identity of some object—as if such a practice needed no justification at all. But my point here is not entirely negative: solutions to the puzzles in question just need to be explicitly and strongly related to more foundational issues if they are to be evaluated in terms of widely accepted theoretical virtues like descriptive and explanatory adequacy.

I have been helped along the way by suggestions and encouragement from a great number of people and I am pleased to have the opportunity to thank them here. Thanks to Aðalsteinn

Hákonarson, Ásgeir Berg Matthíasson, Elisabeth Camp, Robyn Carston, David Chalmers, Bill Child, Lars Dänzer, Dan Dennett, Matti Eklund, Cressida Gaukroger, Geir Þórarinnsson, Peter Godfrey-Smith, Dan Harris, Nat Hansen, Hrafn Ásgeirsson, Marilynn Johnson, Andreas Keller, Krista Lawlor, Ernie Lepore, Eliot Michaelson, Mikael Karlsson, Rachel McKinney, Matt Moss, Nanna Teitsdóttir, Mikel Negugogor, David Pereplyotchik, Ben Phillips, David Plunkett, Michaela Popa, Jim Pryor, Jesse Rappaport, François Recanati, Stephen Schiffer, Barry Smith, Dan Sperber, Robert Stainton, Eric Steinhart, Iakovos Vasiliou, Kathleen Wallace, Deirdre Wilson, Jack Woods, and Dan Zeman.

Special thanks to my adviser, Stephen Neale, who has supported me throughout my graduate studies. By now he has read and commented on innumerable drafts, papers, chapters, and whatnot, always pressing me to do better. Well, I have. I am also very grateful to Michael Devitt, Gary Ostertag, David Rosenthal, and Jesse Prinz, who served on my prospectus committee, and have all provided me with helpful comments and criticism on various aspects of my project. In many ways, Michael has acted as my second doctoral adviser, always ready to talk philosophy and help out, even when we realize that no one else in the packed elevator at the Graduate Center thinks edification is a good companion to asphyxiation.

Dan Harris, Eliot Michaelson, and Ben Phillips have all been particularly assiduous in reading and giving written comments on penultimate or antepenultimate drafts of many of the five chapters that follow. I see now that our discussions have been very important for my overall project at different stages of its development. But, of course, any error is my own responsibility.

I also want to thank my mom and my dad, my sister and my brother, for always supporting me in my endeavours, even if they cannot help looking puzzled when I try to explain what it's all about.

I could never have written any of this if it wasn't for my wife, Nanna, who has been incredibly patient, understanding, and loving. It certainly helps that she is a philosopher of the first rank and when we are not preoccupied with other things she can be mined for suggestions, insight and, of course, knee-jerk reactions to increasingly silly sample sentences. Our wonderful daughter, Þórdís Yrja, was born in September 2012, exactly when I was starting to think

seriously about the dissertation. To my mind, the timing was perfect, since I really needed the change in perspective afforded by fatherhood.

I dedicate this work to the memory of my friend, philosopher, and incredible human being Gunnar Júlíus Guðmundsson, who died far too young. You were philosophy's greatest.

EGU

Reykjavík

18 March 2015

Contents

Introduction	1
1 Confusion is corruptive belief in false identity	7
Introduction	7
1.1 Two models of the mental state of confusion	8
1.1.1 Frege vs. Millikan	8
1.1.2 The toolkit	11
1.2 The Frege model: for and against	13
1.2.1 The objection from unavailable representation	18
1.2.2 Descriptive and explanatory adequacy	23
1.3 Proper functions and malfunctions	26
1.3.1 Proper functions introduced	26
1.3.2 The proper function of singular terms in communication	29
Conclusion	34
2 Puzzle-driven semantics	36
Introduction	36
2.1 What is puzzle-driven semantics?	37
2.2 Three shoddy arguments?	46
2.2.1 Kripke's theory of semantic reference	47
2.2.2 Reimer's argument against intentionalism	51
2.2.3 Perry on 'automatic' indexicals	57
2.3 Donnellan on descriptions: Limiting case?	60
2.4 Interlude on puzzles and anaphora	62

Conclusion	67
3 Arguing for Gricean intentionalism	69
Introduction	69
3.1 Gricean speaker meaning	70
3.2 Argument from underspecification	76
3.2.1 Indexicality	77
3.2.2 Ambiguity	78
3.2.3 Ellipsis	79
3.2.4 Illocutionary force	81
3.2.5 What about so-called ‘eternal’ sentences?	84
3.3 Malapropism: Saying without meaning?	88
3.3.1 Objection to intention-based semantics	89
3.3.2 The misarticulation theory of malapropisms	94
3.3.3 Argument from arbitrariness	96
Conclusion	100
4 Explicating speaker reference	103
Introduction	103
4.1 Edenic reference	104
4.1.1 Confusion corrupts the evidence	109
4.1.2 Puzzles are problematic	111
4.1.3 Edenic reference is part of pragmatic competence	112
4.1.4 Edenic reference is explanatorily basic	114
4.2 Coreferential cognizance	116
4.2.1 Do you know what you’re saying?	118
4.3 The edenic theory and King’s account	120
4.4 Objection: Intentional confusions	122
Conclusion	125

5 Representation without thought	127
Introduction	127
5.1 Blueprints and coreference	128
5.2 Arguing for the monadic thesis	132
5.2.1 Argument from the arbitrariness of language	134
5.2.2 Argument from reflexivity	137
5.2.3 Argument from middle voice	140
5.2.4 Argument from <i>ad hoc</i> concepts	141
5.3 Mismatch between meaning and content	145
5.3.1 Overarticulated constituents	148
Conclusion	151
Bibliography	153

Introduction

A fundamental question in the philosophy of language is, ‘What determines the reference of a singular term on an occasion of utterance?’ I argue that the best answer is that reference is determined by the speaker’s referential intention. Thus, if the speaker’s plan in uttering or inscribing a singular term—say a name or indexical—on a given occasion is to direct the hearer’s or reader’s attention to a particular object *o* then it follows, simply, that *o* is what the speaker refers to on that occasion. Referential intentions are defined in terms of Gricean communicative intentions more generally, i.e. intentions to have specific cognitive effects on the hearer by way, partly, of the hearer’s recognition of that very intention.

Although the answer is simple, intuitive and, most importantly, an integral part of a very successful theory of meaning and communication, it faces significant problems and objections. One kind of problem or consideration is so pervasive in philosophy that it marks its own independent tradition of thought, or so I argue. The problem can be stated as follows. Speakers are often—either momentarily or persistently—confused about the identity of the objects to which they intend to refer but still, even in such cases, it seems perfectly possible that they succeed in referring to one object rather than another. I may think the keys in my hand are in fact my own, when they are not, so I am confused, but it seems like I can still, unproblematically, refer to them with some public language expression, say a demonstrative. Thus, the thought continues, reference occurs despite any internal conflict in the corresponding intention and so it must be determined by something else.¹

This type of objection springs from what I call ‘puzzle-driven’ theorizing. Large swaths of philosophy are driven by puzzles, but none more so, it may seem, than philosophy of language. Since Frege theorists have, implicitly or explicitly, set themselves the task of building coherent

¹For a recent statement of exactly this problem, and its relevance to intention-based theories of meaning and reference, see Jeff Speaks (forthcoming).

systems of description designed specifically to capture as many possible cases of confused identity as humanly conceivable—‘Frege cases,’ ‘Paderewski cases,’ ‘Twin-Earth cases,’ and so on. I take a stab at this tradition, arguing that it is not theoretically motivated and that its dismal success rate is reason for pessimism. The debate has been badly served by an emphasis on the apparent failure of substitutivity in belief-reports—but, as I hope to show, the more basic phenomenon is the mental state of confused identity.² Further, I show in some detail that the objection is wrong-headed and that the intention-based approach to meaning in general and to singular terms in particular stands untouched.

J. L. Austin made a distinction between doing something by mistake and doing something by accident. As I understand it, it is simply a distinction between doing something wrong because of a false belief on the one hand and because of a failure in performance on the other. When I shoot the wrong donkey because I believe, even if just momentarily, that it is identical to the donkey I was supposed to shoot, I do it by mistake. When I take aim at the right donkey but, being a terrible shot, kill the donkey standing next to it, I did the wrong thing by accident. Counterexamples to the Gricean picture of reference, and the intention-based notion of ‘what is said’ more generally, fall into three categories.

First, there are mistakes, where a speaker either believes falsely that a single object is two objects or believes falsely that two objects are single object and utters a singular term intended to refer to ‘that’ object.

Second, there are accidents. This is the category of malapropisms or speech errors more generally. The speaker has some expression as her target but makes a slip along the way and happens to utter some different expression. Theorists have used such examples, just like cases of confused identity, to argue that speakers often refer to objects or make statements which form no part of their communicative intention.

Third, there is pretense. Taking that notion very broadly, speakers often pretend to say things or refer to things, while actually meaning or intending something else. On the Gricean picture, irony, metaphor, overstatement, and figurative speech in general, are explained by

²Jennifer Saul (2007) has, to my mind, already made inroads in establishing this point by suggesting examples where substitutivity fails in sentences containing none of the standard ‘opacity’-producing expressions.

holding that the speaker makes as if to say one thing—i.e. merely pretends to say it—while in fact meaning something quite different. So the speaker doesn't necessarily *say* anything. Many theorists disagree with this description and take it that figurative speech supports the view that one can say and mean something without it being part of any communicative intention. So I can say and mean that the weather is nice by uttering 'Nice weather,' while looking at the blizzard outside, trying to be ironical. It is just that I also mean that the weather is dreadful.

My focus is mostly on the first two categories, so I comment on them first. I argue that both should be treated as pragmatic performance errors in theorizing about the semantics and pragmatics of natural language. In place of puzzle-driven semantics, I propose an 'explanation-driven' approach. Specifically, philosophy of language ought to be driven by the basic task of explaining how it is possible for humans to communicate and express ideas, beliefs, desires so efficiently and successfully by uttering the particular sounds or making the particular inscriptions they do.³ Any approach of this kind—Gricean intention-based or not—can offer compelling reasons for idealizing away from performance errors in its definition of communicative success. Of course, this does not mean that such errors are irrelevant, they just form no part of the primary explanandum of the final theory.

Pretense calls for a separate treatment altogether but, fortunately, the arguments and proposals made here are neutral on the vexed issue of figurative speech. But something akin to pretense comes up twice in the dialectic. First in the context of intentional malaproisms in Chapter 3 and secondly in the context of deliberately obfuscatory speech acts in Chapter 4. In both cases I argue that the cases in question do not provide occasion for disagreement between different explanatory theories.

The positive theory that emerges, called the 'edenic' theory of reference, places cognitive constraints on the proper performance of the speech act of referring with an expression. In Grice's own terms these are 'optimality' conditions on the mental states involved in a given piece of linguistic behavior. There are two such constraints. The first states roughly that singular term utterances fulfill their proper function only when the speaker is free from any corrupting confusion of identity. The second is a constraint on coreference and states, roughly,

³The formulation is based on Stephen Neale's (2004: 71–72, forthcoming).

that in uttering a sentence containing more than one singular term the speaker must either have a coreferential or a noncoreferential intention. I argue, among other things, that the constraints characterize speakers' basic pragmatic or communicative competence.

A word or two on orientation before giving quick overviews of each chapter. The argument and methodology on display here are thoroughly theory-driven and anti-intuitionistic. As custom demands, I discuss a host of examples or cases where the intuitions of theorists are drawn in different, inconsistent directions. Many philosophers take such cases as evidence for or against a theory and argue accordingly. But, I argue, the plausibility of so proceeding varies considerably from case to case and there are particular problems involved when pragmatic performance errors, due to confused identity, are taken as primary data. My approach is, however, to put the theory in the driver's seat and try to accept its predictions in particular cases. And if the cases can be redescribed—without undue contortions—in terms of the theory then all is well as far as that goes. But, I argue, atypical cases of this sort, considered in isolation, do not provide independent or direct evidence for the theory.

I am profoundly pessimistic about theory-neutral, intuition-based arguments in philosophy, so I prefer to state a problem which is clearly worth solving and follow through on the predictions of an explanatory theory which seems to have some prior probability.

In Chapter 1, I develop a new theory of confused identity as a representational mental state. Ruth Millikan has argued that such confusion is (i) an 'error of its own kind,' which (ii) 'corrupts' the basic function of cognition, and (iii) cannot be defined in terms of false beliefs. To illustrate, she asserts that if I thoroughly confuse Bill and Biff and treat them as if they were a single individual, I have a corrupt singular concept of Bill/Biff. The concept tracks two objects while it is 'designed' to track only one. Moreover, the belief that Bill is identical to Biff cannot be attributed to me because then, as she puts it, I "should have to have a thought of Bill and another of Biff, which thoughts I was disposed to coidentify. But a thought of Bill that is other than my thought of Biff is exactly what I do not have" (Millikan 1994: 97).

I show in detail that although (iii) is wrong, (i)-(ii) are correct. Specifically, I show that false identity beliefs must be attributed to confused agents if their linguistic behaviour is to be explained at all. On my alternative view, confused identity is defined as a mental state which

manifests itself in two basic ways. Either the agent implicitly and falsely believes that two objects are identical ('combinatory confusion') or she implicitly and falsely believes that a single object is two objects ('separatory confusion').

In Chapter 2, I put this theory to work to define the tradition of 'puzzle-driven' semantics. I show how this tradition has dominated philosophy of language and mind since Frege. Theorists in this tradition describe contexts in which confused speakers utter singular terms and refer thereby to an object about which they have false identity beliefs. Cases of this kind have been thought to pose various semantic puzzles, such as: What did the speaker, in that context, *actually* refer to? Philosophers wrongly suppose it is an important theoretical task to answer questions of this sort.

Indeed, it is doubtful that semantic claims about confused speakers can be generalized to speakers and hearers in any 'normal' context. By analogy, cognitive scientists don't assume that claims that are true of the mechanism of face-perception in prosopagnosiacs are also true about such mechanisms in normal humans. But many of the most influential arguments in the philosophy of language fail, or so I argue on a case-by-case basis, because the speaker's identity confusion disrupts the evidential proper function of his utterance of a corresponding singular term.

In Chapter 3, I start developing the view that semantics should be explanation-driven and intention-based. I present the general outlook of Griceanism and introduce two arguments in its favor. The first is based on recent research in pragmatics, especially on cases where the intended meaning of the speaker is radically underdetermined by the linguistically encoded meaning of the sentence in the language. The second uses experimental data from phonetics and phonology to show that intentionalists provide clear and convincing explanations of the nature of malapropisms and various types of speech error.

Chapter 4 then proposes, as described above, doxastic constraints on any Gricean notion of the act of referring with a singular term, such that problems arising from conflicting referential intentions are completely avoided. I argue that the constraints—i.e. the 'edenic' and the 'cognizance' constraints—characterize a notion of reference which plays a fundamental explanatory role in the final semantic theory.

Finally, in Chapter 5, I argue that the edenic theory of reference calls for a change in the way we think of the logical form of coreferential constructions. It is natural, on the intention-based view proposed here, to think of the linguistic meaning of sentence types in terms of non-propositional blueprints or templates. Such blueprints encode very general instructions for message construction. In the case of simple sentences containing only one singular term and a predicate, the application of the edenic theory is quite straightforward.

Coreference is different. I present four arguments for thinking that edenic and cognizant utterances of polyadic coreferential sentences really express monadic contents. Uttering the polyadic form communicates a monadic property which is applied to the object referred to, even if it is referred to multiple times in the same clause. This is called the ‘monadic’ thesis. If the monadic thesis is true, and I argue that it is, it gives rise to a new kind of structural mismatch between linguistic meaning and propositional content. I show how intention-based semantics gives a better explanation of this kind of mismatch than other theories, for example minimalism and hidden indexicalism.

Chapter 1

Confusion is corruptive belief in false identity

Introduction

People are often confused about the identity of objects. They think one thing is two things or two things are one. Since Frege argued for the distinction between sense and reference, this type of mistake has been the driving force behind a host of influential arguments and theories in philosophy of language and mind. Until recently, however, theorists did not pay much attention to the metaphysics of confused mental states. The point of this paper is to argue for and present a new theory of confusion, called the ‘Frege model.’ Hammering out the details of such a theory is important for many reasons. For example, it provides the background for a robust evaluation of the puzzle-driven methodology that has dominated philosophy of language and mind for quite some time—where one asks what speakers refer to by some term (‘Hesperus,’ ‘Paderewski,’ ‘water,’ etc.) about which they are confused. But also, there is no consensus among theorists about how identity confusion ought to be described, with some even arguing that it is too perplexing to be construed as a mental state at all.

In §1.1 I describe two plausible models of identity confusions. The first is the Frege model already mentioned and the second is called the ‘Millikan model.’ I then introduce a few tools and distinctions which will be important in adjudicating between the competing models. In §1.2 I explain the Frege model in more detail and present an objection to it—called ‘the argument from unavailable representation’—and show in some detail why it fails. I also argue that the Frege model is superior to its competitor both in terms of descriptive and explanatory adequacy.

Interestingly, however, many of the insights of the Millikan model can, and should, be incorporated into our final theory of confused identity. As Ruth Millikan has argued extensively,

confused identity is a type of corruption of the *proper function* of our capacity to reidentify particulars. In §1.3, I explain how Millikan's notion of proper functions can be applied to the case of confused identity, first in thought and then in communication. Particularly, I show how confusion essentially disrupts the proper function of singular terms in natural language, even if the Frege model is correct.

1.1 Two models of the mental state of confusion

Identity confusion is a mental state of two basic types. *Combinatory confusion* is when an agent takes two things to be a single thing. *Separatory confusion* is when an agent takes a single thing to be two things. In principle an agent can be confused by taking n things to be m things for any distinct natural numbers n and m . For convenience, I'll focus on simpler cases involving only one or two objects.¹ A dog who holds a bone in his mouth and takes its reflection in a pond to be a different bone suffers, albeit momentarily, from separatory confusion. Another dog who chases two similar rabbits, only ever seeing one of them at any given moment, might suffer from combinatory confusion. If the rabbits switch roles behind a rock, appearing to the dog as if a single running rabbit disappears from sight only for a moment, he seems to have some kind of mental representation which is 'supposed' to help him track a single rabbit but 'really' tracks two rabbits.

1.1.1 Frege vs. Millikan

How exactly should identity confusions like those attributed to the two dogs be explained and modeled? There seem to be two plausible ways to go here (see, e.g., Lawlor 2005).

¹Note that this is not supposed to be an analysis of the meaning of the word 'confusion' in English. The characterization is an attempt to describe what confusion itself consists in. It seems like the *word* 'confusion' is partly negative and is not properly applied, in English, unless the agent is taken to be somewhat responsible for her sorry epistemic state. This would arbitrarily disqualify many cases that a general theory of confusion ought to capture. Further, theorists who have written specifically about confusion tend to think only of combinatory confusion (cf. Camp 2002; Lawlor 2007). This may simply reflect common beliefs about the etymology and morphology of the lexical item 'confusion' in English. Translations into other languages do not always give rise to the same tendency (German 'Verwirrung,' Icelandic 'ruglingur,' etc.).

1. *Frege model.* According to this model confusion is a mental state that consists in someone's having identity beliefs that are false. It is on this basis that Fregeans have postulated *modes of presentation* or *senses* as semantic values for linguistic or mental representations. Some theorists think of modes of presentation as identifying conditions specified by descriptions. Most famously, Frege proposed that someone might think of Venus via the identifying condition of *being the morning star* or of *being the evening star*. And it is not necessary that they realize that these conditions determine the very same object. Modes of presentation, however, are an optional feature of the Frege model. It is 'Fregean' only because of its role in some of the most influential arguments for postulating such modes.

A rough formulation of the Frege model is as follows. Combinatory confusion consists in the fact that an agent *A* explicitly stands in some propositional attitude relation to distinct objects *a* and *b*, and believes that $a = b$. I emphasize that the Frege model of confusion does *not* presuppose a Fregean view of propositional content. Rather, Frege, and many others, have seen confusion cases as a reason for Fregeanism about content. And, certainly, *A*'s mental or linguistic representations of *a* and *b* will be indistinguishable from *A*'s perspective. Separatory confusion consists in *A* explicitly standing in some propositional attitude relation to *a* and *b* where $a = b$, while *A* believes that $a \neq b$. In this case, of course, it appears to *A* as if her representations concern two different objects. The Frege model is stated more precisely in §1.2.

2. *Millikan model.* Ruth Millikan (1994, 1997, 2000) develops a subtle theory of confused thoughts. On her view, identity confusion cannot consist in having a false belief about an identity, rather, it is an error of its own kind (1994: 96, 2000: 173). In the "central cases," misidentifying the object of one's thought is "an act that muddies the thought involved, corrupting the inner representational system" (1994: 75). This is taken to mean that confused representations involve concepts which are themselves confused. Here is how I propose, in more detail, to flesh out the Millikan model. A confused concept is one where the cardinality of the set of objects in its actual extension differs from the cardinality the set is, in effect, taken to have by its possessor. Consider the dog chasing

the two rabbits. He, let us suppose, possesses a singular concept or quasi-concept C_R such that it actually refers to a spatially discontinuous and gerrymandered object that consists of two individual rabbits, Rabbit *A* and Rabbit *B*.² The concept C_R is corrupt, however, since its proper function is to refer to a single spatially continuous object, specifically *the rabbit being chased*, and, one may add, the dog seems to quasi-believe that C_R really performs this kind of function.³

Now consider separatory confusion. Suppose that the dog carrying the bone in his mouth has a singular concept or quasi-concept C_B of that particular bone. On the Millikan model, as soon as the dog confuses the reflection of the C_B -bone in the pond for some *different* bone, his tokening of the C_B -concept refers to nothing. The actual reference of the C_B -concept, then, is a non-existent object which consists of *that-bone-in-the-mouth* without consisting of *that-pond-reflected-bone* at the same time. But an object that simultaneously consists of itself and does not consist of itself is impossible. Thus the extension of the C_B -concept is the empty set, although its proper function is to have a single spatially continuous object in its extension.

Two points of clarification. First, the models are supposed to capture the *metaphysics* of confusion, i.e. what the mental state of confusion consists in, at some level of abstraction. We will not be concerned with the *epistemology* of confusion here, i.e. the question of how one ought to interpret the beliefs and utterances of confused agents (on that question, see Camp 2002; Lawlor 2007). Clearly, answering the metaphysical question will have consequences for how one deals with the epistemological question, but they should be kept apart at the outset.⁴

Second, some theorists have argued that identity confusion cannot be a mental state at all—but usually they only have in mind combinatory confusions. Joseph Camp (*ibid.*) and Millikan agree, against the Frege model, that confusion cannot consist in false beliefs.⁵ But Camp goes

²Or it refers partially to Rabbit *A* and partially to Rabbit *B* at the same time (cf. Field 1973). As we will see, the arguments of this Chapter go through on either assumption.

³The notion of ‘proper function’ is explained in §1.3 below.

⁴For example, it seems like supervenience theories of confusion are much more plausible as an answer to the epistemological question than the metaphysical one (cf. Field 1973; Lawlor 2007). On the importance of keeping epistemology and metaphysics separate, see Devitt (2010).

⁵Relatedly, many have argued that *clinical delusions*, and similar afflictions, cannot consist in having a certain specific set of false beliefs. But Lisa Bortolotti (2010) provides a compelling

further and argues that confusion is not really a mental state. Ruth Marcus (1981, 1983, 1990) toes the same line when she argues that it is impossible to believe what is impossible. She has in mind both combinatory and separatory confusions (and more, it would seem). Marcus claims that it only appears to us that agents have confused identity thoughts because our notion of belief is too linguistically oriented. Agents can certainly utter *sentences* that describe impossible states of affairs but, on Marcus' view, belief is dispositional and not linguistic. If an agent *really* believed something impossible it should make her behavior aberrant in some relevant way (for a compelling counterargument, see Richard 2013). I will assume, if only because externalism about 'mental states' might possibly be true, that confusion should be thought of in strictly psychological terms and stick to the two more positive models.

1.1.2 The toolkit

How exactly should one adjudicate between the Fregean and the Millikanian models? We will need a few theoretical tools. More specifically, we need two criteria of theory evaluation and two distinctions. Let us look at the criteria of evaluation first.

1. *Descriptive adequacy.* As Marcus (1983) emphasizes, the theory must apply to both linguistic and non-linguistic agents. Confusion does not require much in the way of intellectual capacities.
2. *Explanatory adequacy.* Our theory of confusion needs to be explanatorily basic. The more parsimonious model should be preferred over the other.

The criteria may interact in somewhat predictable ways. For example, if theory T1 applies to a larger set of cognitive agents than T2 there is reason to suppose that T1 is also explanatorily prior. This might be because the cognitive system posited by T1 is shared by all confusion-susceptible agents, making the explanation more parsimonious and less intellectualized than the one offered by T2. One might be tempted to suppose, for instance, that the Millikan model carries the day on both (1) and (2) because it avoids positing an intellectually sophisticated counterargument, showing that even such cases are best described in terms of a network of implicit and explicit beliefs.

capacity for having thoughts involving identity. But, as I will soon argue, this not right and the issue requires a more thorough treatment.

Secondly, we must have a reasonably clear distinction between explicit and implicit propositional attitudes. To fix ideas, I propose to assume a bare bones theory of beliefs and other propositional attitudes as mental representation: there are mental states like thoughts, beliefs, and desires and there exists some internal representational system in which these are couched. If one explicitly believes at time t that everyone likes to party then some corresponding mental representation is tokened in one's internal 'belief box' at t . The mental representation, then, has a content that is specified by a *that*-clause. Call the contents of mental (and public language) representations 'propositions.' Sometimes I talk as if agents believe propositions directly, but this is only shorthand for the more cumbersome locution with mental representations as intermediaries. The notion of a 'proposition,' for our purposes, is only a classificatory notion: propositions are not mental states or mental representations but, as Perry (2012: 27) puts it, "abstract objects that we use to classify [such] states and events by the requirements their truth (or some other form of success) impose on the rest of the of the world." Mental representations and token sentences in a natural language, however, are concrete events with causes and effects.⁶

Now, this background provides for at least two notions of 'implicit' attitudes. First, there are *derived* attitudes or beliefs. These are the beliefs that can be swiftly derived from the set of one's explicit beliefs. Thus, if one believes explicitly that the number of planets is 8, one thereby believes implicitly that the number of planets is lower than 50. This latter belief need not be represented in one's belief box, but can be derived fairly easily. Note that this does not give us a precise boundary where explicit belief stops and implicit belief starts. Secondly, there are *governing* or *guiding* propositional attitudes. These are implicit representations or rules (or biases) in accordance with which mental processes move from one explicit representation to another. If there are psychological laws they may belong here, and clearly such rules or

⁶Cf. Churchland (1979: 105); Dennett (1982: 123–126); Dresner (2006); Fodor (1978: 200–202); Moltmann (2003: 111–113); Perry (2012: 26–31); Stalnaker (1984: 7–10). For a detailed treatment, see Matthews (2007).

laws can, but need not, be explicitly represented in an agent's belief box.⁷ Indeed, it may be best to think of them as merely embodied rules rather than actual representations (Devitt 2006: §3.1). Either way, they are rule-like entities attributed to speakers to explain a wide variety of intelligent behavior.

Finally, we need to distinguish between the language, or other system of mental representation, used by the confused agent and the language used by the theorist in describing and explaining the confused mental state. For simplicity, call the theorists' idiolect 't-language' and the confused agent's idiolect 'c-language.' The c-language contains singular terms that correspond to mental representations which either (i) mentally combine two objects into one or, together with another representation, (ii) mentally separate one object into two. Accordingly, we can call such singular terms in the c-language *confused* (or 'c-terms' for short). The t-language cannot contain any confused singular terms or, at least, it cannot contain the very same confused terms as the c-language under consideration. Otherwise the t-speaker will be assumed, from the start, to believe falsehoods or have corrupt thoughts about the subject matter of her theory.

For now, I take this distinction to be fairly intuitive and hold off a detailed account until the argument calls for it. In the next section, the Frege model is stated more precisely and explained with examples. Then I formulate an influential objection to the model and argue that the objection fails.

1.2 The Frege model: for and against

Now let us state the Frege model in detail. The basic definition is, as expected, given entirely in the t-language. The agent's own c-language comes into play later.

⁷Dennett (1978: ch. 6); Fodor (1985, 1987). Note that the argument in this paper does not essentially depend on a particular theory of propositional attitudes. Dispositionalism (e.g. Schwitzgebel 2013), instrumentalist or interpretationist views (e.g. Dennett *ibid.*) are completely compatible with what follows. As long as the explicit-implicit distinction can be drawn in a reasonably clear manner, we can safely set sail. Furthermore, Peter Godfrey-Smith's (2004, 2005) model-cum-construal theory of folk psychology also fits my general agenda perfectly. He argues for the legitimacy of both realist and more interpretationist construals of mental states like beliefs and desires.

Frege model of confusion

Agent A suffers from identity confusion at time t iff for any propositional attitudes V and W that A explicitly holds at t , (i) $A \text{ Vs } Fa$, (ii) $A \text{ Ws } Gb$, and only one of (iii–iv) is true; (iii) $a = b$ and A believes that $a \neq b$ (or lacks the belief that $a = b$), (iv) $a \neq b$ and A believes that $a = b$ (or lacks the belief that $a \neq b$).

Assume that ‘ a ’ and ‘ b ’ are arbitrarily chosen singular terms in the t -language and ‘ F ’ and ‘ G ’ are arbitrary predicate expressions of the same language.⁸ Note that in the definition the properties represented by ‘ F ’ and ‘ G ’ may be any property A can cognitively represent. They can, for example, be identical or mutually exclusive. Similarly, V and W can be, for example, the same (e.g. belief), completely different (e.g. hope and fear respectively), or mutually exclusive (e.g. belief and disbelief respectively). Finally, the definition employs the terminology of explicit propositional attitudes ‘at a time’. What this means is just that there is some specific time span in which A has two explicit mental states whose contents represent a and b . In many cases, the two states can be collapsed into a single mental state whose content represents a relation holding between a and b , i.e. (i–ii) can be: $A \text{ Vs } Rab$. But the terminology is used here, firstly, to exclude cases where the mental states occur at two very distant points in time and, secondly, to contrast them with the identity beliefs (iii–iv) that A may have implicitly.

The Frege model is intended to describe a particular type of mental state which has, for some reason, figured quite prominently in philosophical arguments about meaning and content. But the model is much more general than the examples of the philosophers would suggest, i.e. it seems to cover more types of cases, and this feature needs to be briefly explained.

To this end, let’s focus on cases of separatory confusion. Here is a typical scenario in which an agent (call her ‘Lois Lane #1’) satisfies the above definition of confusion. Lois #1 believes that Superman can fly. She also believes that Clark Kent cannot fly. As a matter of fact (assuming this is fact rather than fiction), Clark Kent is Superman. And, finally, Lois #1 does not believe that Clark Kent is Superman. Now consider a case that is less ‘typical’ (relative to the massive

⁸By ‘singular term’ I mean any expression that the t -speaker intends to stand for an object, it need not be a referring expression; it can be, e.g., a proper name, demonstrative, indexical, singular possessive construction, or a definite description used referentially (using an expression ‘referentially’ is explained in Chapter 4, p. 106, with the definition of ‘Gricean speaker w -reference’).

literature on paradigm confusion-cases). Another Lois—‘Lois #2’—is likewise confused, on the proposed definition, if she hopes that Superman saves her and wonders whether Clark Kent has read Shakespeare while not believing that Superman is Clark Kent.⁹

Many philosophers have argued that someone like Lois #1 is confused but not *irrational*. On one account of directly referential singular terms, however, it would follow that Lois #1 is being irrational. Let us call this account ‘Millianism.’ According to Millianism the content of a belief expressed by referring directly to object *o* will have an *o*-dependent truth condition.¹⁰ Sometimes this is explained by saying that the content, or ‘semantic content,’ of a singular term when uttered on some occasion, as part of a simple sentence, is exhausted by its referent. Then it is said that the proposition expressed ‘contains’ the referent itself, rather than any identifying condition on or property of the object in question.¹¹ In any case, Millianism would appear to imply that Lois #1 is irrational since she stands in the belief relation to two mental representations the contents of which are contradictory propositions containing the very same object *o*. If the semantic contents of ‘Superman’ and ‘Clark Kent’ are identical Lois #1 should be able to discover this by using nothing but logic and reason. This has lead many philosophers, since Frege, to argue for a two-tiered semantics in which different singular terms are associated with different modes of presentation. This is supposed to explain how someone like Lois #1 can actually hold these contradictory beliefs without really being irrational. On a Fregean view, no amount of mere introspection and logical acumen can help her discover that she believes two contradictory propositions about a single individual. The propositions contain senses or modes of presentation of that individual and Lois #1 just does not know that these are modes of presentation of one rather than two individuals.

⁹This formulation does not (controversially) assume that *whether*-clauses and *that*-clauses refer to the same (type of) entity. Nevertheless, it is reasonable to think that the phenomenon identified by the definition of confusion will apply in both types of cases, since both may involve reference to objects. See Hanks (2007) for discussion.

¹⁰A proposition *p* has an object-dependent truth condition iff for an arbitrary possible world *w*, $\exists x(p$ is true w.r.t. *w* iff in *w*, . . . *x* . . .) (Recanati 2012: 15). We will assume, although this is far from trivial, that this definition can be extended to satisfaction conditions other than truth conditions.

¹¹On Millianism or the theory of ‘direct reference,’ see, e.g., Abbott (2010); Bach (1987); Devitt (1981a, 1989, 1996, 2012b); Donnellan (1966); Kaplan (1989b); Kripke (1977, 1980); Marcus (1961); Neale (2008); Recanati (1993); Richard (1983); Salmon (1986a, 1989); Saul (1997, 1998); Schiffer (1978); Soames (1989a, 2002).

What about Lois #2? She is confused about the identity of Superman/Clark Kent yet such examples have not been used to motivate Fregean theories of content. It can be shown, however, that this case gives similar grounds for Fregean theories of content. Thus the definition is generalized to any agent bearing arbitrary attitudes to (Millian) propositions containing the referents of both a and b while not standing in the belief relation to the true proposition *that* $a = b$.

Here is a way of arguing for this. At time t , let us assume, Lois #2 tokens the hope that Superman saves her and the wondering-whether Clark Kent reads Shakespeare. But it is not the case at t that Lois #2 believes explicitly that

- (1) $\lambda x \langle$ she tokened the hope that x saves her and she tokened the wonder-ing-whether x reads Shakespeare $\rangle o$.

Where $o = \text{Superman} = \text{Clark Kent}$. Still, this very belief can be *derived* from the set of her beliefs at t and is then, arguably, one type of implicit attitude. The derivation takes Millianism for granted and, plausibly, assumes that conscious, reflective individuals can reliably form true higher-order beliefs about their own propositional attitudes. Using only variables, at t , Lois #2 tokens a V -ing of Fa and a W -ing of Gb . Assuming that she knows some symbolic logic, she can then always derive the proposition that, at t , she believes that $\lambda x [\lambda y \langle$ she tokened a V -ing of Fx and a W -ing of $Gy \rangle a] b$. And since in this case it is true *that* $a = b$ the Millian theorist must predict that there is only a single object o to which Lois #2 refers in deriving the λ -sentence. Thus, given these assumptions, the logician Lois #2 must believe (1) or something like it.

Lois #2 does not explicitly believe the negation of (1) at t , yet this can also be derived from her beliefs at t . To make the point briefly, if one were to put (1) to her in the form of a question, she would respond in disbelief if she responds honestly and in accordance with her actual beliefs. Here is the question I have in mind: ‘Do you believe that having the property of being an x such that you have a token hope that x saves you and you have a token wonder whether x reads Shakespeare applies to a single object o , o being either the object to which you refer by ‘Superman’ or the object to which you refer by ‘Clark Kent’?’ Again, assuming that she knows some logic, Lois #2 ought to understand such contorted questions perfectly. It seems,

then, that Lois #2 implicitly believes a contradiction.

It is my contention that if a two-tiered Fregean semantics can be motivated by taking Lois #1-type cases as data it should be equally well motivated by Lois #2-type cases. This contention is one reason for making the Frege model of confusion so general. The definition is the most general formulation of the reason many theorists believe in Fregean senses or something playing the role of such senses. Here is a simple argument for this. If it is right that Lois #2 has inconsistent beliefs then this would be avoided as soon as one adopted a Fregean theory of content. According to a typical theory of this sort the proposition believed by Lois #2 can only contain modes of presentation (call them 'MPs'), never the objects themselves. If she wonders whether Clark Kent reads Shakespeare she stands in the wondering relation to (a mental representation the content of which is) something like the following proposition:

$$(2) \quad \langle MP_C, MP_S \rangle$$

There are two components in this proposition: Lois' mode of presentation of Clark Kent (' MP_C ') and her mode of presentation of the property of reading Shakespeare (' MP_S ').

Ultimately, the derivation of (1) fails on Fregean assumptions because there are only two modes of presentation that Lois #2 can possibly associate with the occurrence of 'o' at the end of (1). On a Fregean picture of content, she must understand 'o' in one of these two ways. First, there is the MP she associates with the name 'Clark Kent.' At t , however, Lois #2 did not token a hope the content of which had her mode of presentation of Clark Kent as a component, thus this cannot be the MP associated with 'o' in (1). Second, there is the 'Superman'-MP. At t , Lois #2 did not token a wondering whether the content of which had the 'Superman'-MP as a component, thus this cannot be the MP associated with 'o' in (1). Therefore, there is no Fregean interpretation of the λ -expression in (1) on which the claim made by it can be true and thus (1) cannot be derived. Therefore, Fregeanism saves Lois #2 from irrationality because only the negation of (1) can actually be derived from the set of her beliefs at t .

In sum, the case of Lois #2 gives an equally strong reason to accept Fregean MPs as does the more traditional case of Lois #1. The alleged motivation provided by the traditional cases is based on the idea that, on a Millian theory of propositions, the confused agent is also irrational. The agent should, by using reason alone, be able to find out that she is confused. And this

unwanted prediction follows in the other type of case just as well. On Millian assumptions, it seems like reason alone should suffice for Lois #2 to discover that she is confused. Thus the definition of the kind of identity confusion philosophers have been interested in should cover both types of cases and this is exactly what the Frege model accomplishes.

1.2.1 The objection from unavailable representation

Camp (2002: 33) and Millikan (1994: 97) have voiced similar arguments against something like the Frege model, claiming that the relevant representations of identity are unavailable to the confused thinker. In this section, I reconstruct this objection, and refute it. Then I turn to the criteria of descriptive and explanatory adequacy.

The objection is clearer when stated in terms of combinatory confusion. Millikan's (ibid.) uses the example of Bill and Biff. Imagine they are identical twins and *A*, their acquaintance, has met both of them on multiple occasions. But *A* has never seen them both *at the same time* and has in fact enjoyed Bill's company for exactly the same amount of time he has enjoyed Biff's. Neither person commands a dominant role over the other in *A*'s mental economy. To make things easier for us (the theorists), assume this is some sort of ploy, and the twins have intentionally led *A* to believe that they're a single person called 'Phil.'

Thus, if we go along with the popular 'mental file' metaphor, *A* puts all the information gathered from any epistemic encounter with Bill or Biff into the very same mental file, labelled with a single 'tag'—represented linguistically as 'Phil.'¹² Thus I will assume that *A*'s c-language contains no more than one singular term or singular mental representation of Bill/Biff, because all the relevant information is contained in but a single mental 'repository.' According to the objection from unavailable representation this scenario makes it impossible, in any reasonable sense, for *A* to have two distinct representations, one of Bill and the other of Biff. *A* has only one relevant representational item at her disposal, i.e. 'Phil.' But then the Frege model cannot be right, as it clearly requires this possibility.

Camp and Millikan state the objection, then, as a *reductio* of (3).

¹²See, e.g. Lawlor (2001); Millikan (2000); Perry (2012); Recanati (2012). It is not to my purpose to evaluate the theory of mental files as such here. The framework is part and parcel of the objection, as I understand it, and it is assumed here for sake of argument.

- (3) *A* believes that Bill = Biff.

But notice that, strictly speaking, the Frege model is not committed to (3). The proponent of the model can insist on the weaker version of (3), namely: *A* does *not* believe that Bill \neq Biff. In that case, the model is only committed to the agent *lacking* a certain true belief. As should be clear, however, this point does not help the proponent of the model here since it (the model) still requires the confused agent to have two distinct representations that, in effect, distinguish Bill and Biff. This is exactly what Camp and Millikan deny. Furthermore, it seems like the strong version (where implicit beliefs like (3) are postulated) and the weak version (with mere lack of belief) are roughly equivalent. Both types of state are postulated to explain some aspect of an agent's behavior, mental or otherwise. I suspect that this is merely a verbal dispute. But the formulation is left intact for those who disagree. At any rate, I believe that even the stronger version can be convincingly defended against objections and thus I propose to defend that formulation here.

Let us proceed with the *reductio* then. In *A*'s *c*-idiolect there is one relevant singular term 'Phil_{*c*}' and, according to Camp and Millikan, there are only three possible ways of assigning reference to it. It must refer only to Bill or only to Biff or to Bill/Biff. Double brackets '[[]]' represent functions from expressions to their referents or semantic values.

- (4) [[Phil_{*c*}]] = Bill
 (5) [[Phil_{*c*}]] = Biff
 (6) [[Phil_{*c*}]] = Bill/Biff

The last interpretation (6) is construed, according to the Millikan model, such that Bill/Biff is a spatially discontinuous object consisting of two distinct objects, Bill and Biff. But, according to the objection from unavailable representation, none of the interpretations in (4)–(6) can make (3) come out true. If (4) is true, then how does *A* manage to think a thought 'of Biff'? As Camp would put it, if (4) is true then *A* "cannot think anything at all" of Biff and, therefore, cannot believe that Bill = Biff. *A* could only represent the belief *that* Phil_{*c*} = Phil_{*c*}, but, assuming (4) is true, that is clearly not the same as believing *that* Bill = Biff. Same applies *mutatis mutandis* when (5) is assumed.

But what about (6)? Well, according to Millikan this is similarly problematic. For (3) to be true on this assumption, *A* “should have to have a thought of Bill and another of Biff, which thoughts [*A*] was disposed to coidentify. But a thought of Bill that is other than [*A*’s] thought of Biff is exactly what [*A* does not] have” (Millikan 1994: 97). Therefore, again, (3) cannot be true and the Frege model is demonstrably false.

Camp and Millikan both conclude that combinatory confusion cannot consist in having a false identity belief. But ultimately the objection does not work. There are two reasonable responses to it. First, it is certainly conceivable that the singular term ‘Phil_c’ is simply ambiguous. Some tokenings are about Bill and some tokenings are about Biff (and perhaps some are also about the Bill/Biff amalgam). The ambiguity is, surely, lost on *A* herself, but others might become aware that in her *c*-idiolect ‘Phil_c’ sometimes refers to Bill and sometimes to Biff. Then it is, at the very least, possible for *A* to believe falsely, and *explicitly*, that Bill = Biff.

Neither Camp nor Millikan take this option at all seriously. But this is of course a feature of the popular mental files model of singular cognition: each individual object is causally connected to a singular file or concept—and any confusion leads to a kind of corruption. The linguistic item in question is tied to that singular mental file. Still, once the distinction between *c*-languages and *t*-languages has been made clear it seems perfectly reasonable to suppose that combinatorily confused singular terms are ambiguous since they are always paired with (at least) two unconfused terms in the *t*-language (more on this presently). As promised, however, I will assume that ‘Phil_c’ cannot be ambiguous.¹³

The concession poses no problem, however. The objection from unavailable representation can be countered more powerfully without making any assumptions at all about the actual content of a confused singular term like ‘Phil_c.’ The question, What is the actual referent of confused singular terms? is not as fundamental as Camp and Millikan seem to think. Confused identity can be captured and defined without committing to a view on that thorny issue. Armed with the Frege model and the distinction between *c*-language and *t*-language, this is exactly what I propose to do.

¹³Indeed, if ‘Phil_c’ is an item in some mental representational system then, or so some have argued, it cannot be ambiguous in the way described. Thoughts do not appear to be ambiguous in the way natural language sentences so appear (see, e.g., Fodor 1978: 198–200).

Let us start by making the distinction itself more precise. This can be done by introducing two methodological constraints into the dialectic. First, suppose that for any confused singular term a_c in the c-language, a_c can only be mentioned and never used in the theorists' t-language. Otherwise, we will automatically assume that the theorist is confused. That is clearly to be avoided. Second, for any confused singular term a_c in the c-language, the t-language must contain unconfused counterparts. An 'unconfused counterpart' is simply a corresponding singular term in an idiolect of an unconfused speaker. In combinatory confusion, c-term a_c is paired with two t-terms, a_t and b_t : $\langle a_c, \langle a_t, b_t \rangle \rangle$. Normally, in separatory confusion, two c-terms, a_c and b_c , are paired with two unconfused t-terms, a_t and b_t , thus: $\langle \langle a_c, b_c \rangle, \langle a_t, b_t \rangle \rangle$. (Think of the 'Cicero'/'Tully' example here.)

If such expression-pairs occur in the wild, *all* of them may turn out to be homophones. In Kripke's (1979: 153) example of separatory confusion, the c-speaker mistakes Paderewski for two distinct individuals, calling each 'Paderewski.' This c-language fragment then contains two confused names 'Paderewski_{c1}' and 'Paderewski_{c2}' paired with only a single unconfused counterpart—a special feature of homophonic separatory confusions—'Paderewski_t' in the t-language. The inverse Paderewski case (cf. Recanati 2012: 141–142) is a case of combinatory confusion involving homophones. Imagine another c-speaker who takes two individuals, P1 and P2, to be one and both happen to be named 'Paderewski.' Here we must assume that neither P1 nor P2 can lay a claim to being the 'dominant causal source' of the c-speaker's practice of uttering the name or of the attendant mental representation (cf. Lawlor 2007: 162). That is, P1 and P2 are equally responsible for causing the relevant representations of the c-speaker. In this case the c-language contains a single name, 'Paderewski_c' which is paired with two unconfused counterparts, 'Paderewski_{t1}' and 'Paderewski_{t2}' in the t-language. I will try to avoid homophonic examples in what follows, for sake of sanity and all that is good and holy.

Now the argument from unavailable representation can be properly refuted. In cases of confusion the relevant singular representations of the agent must be bracketed as scrambled and corrupt. They cannot be incorporated directly into the t-language. But as theorists, we still want to be able to say something to explain the agent's behavior, communicative or otherwise. And, unsurprisingly, attributing false identity beliefs to agents who may themselves be quite

unable to represent those beliefs explicitly will do the explanatory work required. Thus, when *A* combinatorily confuses Bill and Biff, as shown in her disposition to make certain inferences and assumptions in particular contexts, the theorist ought to explain *A*'s behavior by attributing to *A* an *implicit* identity belief that happens to be false. For instance, *A* will have a complex variety of dispositions to think and express false identities where the identity sign is flanked by two demonstratives 'this' and 'that.' One of them may refer to an object of current perception while the other refers to an object represented by a memory image.

One can easily allow, then, that the confused agent could indeed have thoughts that are 'of Bill' and thoughts that are of 'of Biff.' Remember, also that the agent's 'Phil_c'-representations have two unconfused counterparts in the *t*-language, 'Phil_t' and 'Biff_t.' And these representations are the perfect fit for the theorist who wishes to explain *A*'s 'Bill_c'-related behavior. Echoing what Stuart Hampshire said in a similar context, someone may be unable to explicitly think that *p*, while their behavior can only be explained by the hypothesis that they believe that *p*, given that it is known that they believe that *q*.¹⁴

In terms of the distinction between *derivative* and *guiding* implicit attitudes from §1.1.2, confused beliefs are of the latter kind. The agent is clearly unable to swiftly derive the confused belief from her set of explicit beliefs. But the belief is required to explain and predict her actions. Consider a simple example. Suppose someone (*A*) confuses Bill and Biff and uses 'Bill' for both of them. *A* meets Biff_t and says, 'Hi Bill_c.' How do the folk explain why she uttered the wrong name? Well, simply by saying that *A* believes that *that man in front of her*, i.e. Biff_t, is identical to Bill_t. People give such explanations all the time, and this practice does not appear suspect.

The Camp-Millikan argument could only be evidence for the view that confused speakers can have no *explicit* beliefs of the form 'X is Y,' but according to the Frege model the belief in question can be either explicit or implicit.¹⁵ The point can be made by saying that, since

¹⁴For the Hampshire quote, see Dennett (1982: 164n16).

¹⁵Camp (ibid., 31) advances yet another argument against the notion that confusion consists in false belief. But this argument assumes that the beliefs in question are *general* rather than *singular*. Thus it does not apply to the Frege model. Camp seems indeed to be right that identity confusion cannot consist simply in holding false identity beliefs whose expression is, e.g., an identity sign flanked by two definite descriptions. If nothing satisfies either description, for example, the belief is surely false without the agent necessarily being confused. If one believes, for example, *that the king of France is the economic advisor of Atlantis* one need not be confused

the theorist should not be allowed to simply adopt the language of the speaker in question, one is compelled to endorse a kind of local interpretationism about confused beliefs.¹⁶ There is no need to embrace interpretationism *tout court*, as other unconfused singular terms can be assumed to be shared between the speaker and the theorist. And the theorist can, then, posit explicit internal representations and mental states. In that case, the theorist does not automatically commit to any errors by using the same singular representations as the one who is confused.

1.2.2 Descriptive and explanatory adequacy

As noted above, if a theory of confused belief is to be descriptively adequate it must make room for the fact that nonlinguistic animals and prelinguistic infants are just as confusion-susceptible as other more intellectually sophisticated creatures. Despite appearances to the contrary, the Millikan model holds no advantage in this respect. Using the t-language, the Fregean theorist can easily assign implicit identity beliefs to nonlinguistic agents if such are required to explain their behavior. Camp, Marcus, and Millikan seem to be driven, in different ways, by the intuition that confusion-susceptibility is cognitively more basic than the ability to think thoughts about or involving identity. If thoughts can be implicit the intuition is mistaken and even goes against some of their other claims. For example, according to Millikan's theory the "... central job of cognition is the ... task of reidentifying individuals, properties, kinds, and so forth, through diverse media and under diverse conditions" (2000: xi). But how can reidentification be explained without attributing implicit identity beliefs? I don't see how it can. Millikan suggests, for example, that reidentification is explained by the capacity to recognize when two thought tokens are thoughts of the same. Surely, this formulation does not imply that the agent must explicitly represent *that the subject of one thought token is identical to the subject of another thought token*. But it cannot be denied that the mental process in question is partly explained by assuming that the agent implicitly believes that the subjects are identical.¹⁷

about identity.

¹⁶Cf. Dennett (1982, 1991); Davidson (1984).

¹⁷As John Campbell (1987, 2002: 97–101) likes to put it, when one identifies *o* at *t* and reidentifies *o* at a later time *t'* one must at least be 'trading on' the identity of *o*. But this is just fancy terminology for implicit or tacit belief of some sort. Relatedly, Tyler Burge (2010: e.g. 286–287,

To borrow Fodor's (1985: 24) example, let us assume some type of associationism about mental processes. Imagine, further, that there is a 'principle of association by proximity' in virtue of which thoughts of salt are usually associated with thoughts of pepper. The principle is then, as mentioned, a guiding or law-like attitude which explains why salt-thoughts and pepper-thoughts are invariably linked in trains of thought (see §1.1.2). But the principle itself need not be explicitly represented in the mind. According to standard representational theories of the mind, however, the condiment-thoughts themselves must be explicitly represented.

Something similar happens when a train of thought contains different thoughts that are supposed to be about the very same object. This can occur, for example, in inferences. When a thinker infers *that Cicero is an orator* from the belief *that Cicero is tall and an orator*, she assumes that she is thinking about the same object twice. Although there is no need to assume that she explicitly believes or entertains the thought *that Cicero = Cicero* it helps to explain her inferential behavior if we assume that she believes this implicitly. And if—as Millikan asserts in one of her discussions (1994: 97)—we are not allowed to attribute any such belief to her, explicit or implicit, it is unclear how we are to explain this piece of behavior at all.

Thus the Frege model surpasses the Millikan model in explanatory power, if indeed the latter eschews implicit identity beliefs altogether. It seems reasonable to think, however, that no model of confusion can genuinely avoid postulating such beliefs. The reason for this is the distinction between t-languages and c-languages. If the theorist is to remain unconfused while explaining the behavior, inferential or otherwise, of a confused agent, she can only ever *mention* confused terms from the c-language. But if this requirement is accepted, and I see no reason to why it should not, the Millikan model must postulate false identity beliefs as well. Obeying our modest methodological strictures, here is how the Millikan model can, I propose, be stated more precisely.

Millikan model of confusion

Speaker *A* suffers from identity confusion at time *t* iff for any propositional attitudes *V* and *W* that *A* explicitly holds at *t*, either (1) or (2) is true:

460) argues extensively that creatures need not be capable of *thinking* the 'criteria for reidentification' in order to reidentify individuals and objects. Mere 'perceptual tracking' counts as reidentification. As I'm construing implicit belief, Burge's view and mine are compatible.

1. (i) A Vs F (the referent of ' a_c ' in A 's c-idiolect), (ii) A Ws G (the referent of ' b_c ' in A 's c-idiolect), and (iii) $a_t = b_t$ but A 's ' a_c '/' b_c '-representations presuppose that $a_t \neq b_t$.
2. (i) A Vs F (the referent of ' a_c ' in A 's c-idiolect), and (ii) $a_t \neq b_t$ but A 's ' a_c '-representations presuppose that $a_t = b_t$.

In (1) the speaker suffers from separatory confusion and ' a_c ' and ' b_c ' are paired with corresponding terms in the t-language. For instance, 'Superman' in the c-language is paired with 'Superman' in the t-language, and so on. (2) is combinatory confusion and ' a_c ' is paired with two corresponding terms in the t-language. If A uses ' $Bill_c$ ' to refer to what the t-speaker refers with both ' $Bill_t$ ' and ' $Biff_t$ ' these are paired together. Once the distinction between the c-language and t-language is made there is no reason to keep to the terminology that was introduced before in explaining the Millikan model. That is, we need not say, in (1), that the c-speaker refers to a spatially discontinuous object consisting of two distinct objects. Nor that, in (2), the c-speaker purports to refer to a nonexistent object that consists of itself and not itself at the same time. Millikan's model aimed at bringing in the actual content of items in the confused speaker's inner representational system. But this is not necessary, as the mental state of confusion can be defined without mentioning those contents. This is good since the model need not take a stand on particular content-assignments in these puzzling cases.

But what does it mean to say that representations in the c-language 'presuppose that $a_t = b_t$ or that $a_t \neq b_t$ '? Roughly, that the way in which A uses these representations is best explained by assuming that A implicitly believes one or the other. The representations 'function as if' they were representing a single object while they actually represent two objects. Or the representations 'function as if' they were representing two objects while they actually represent only a single object. In other words, the relevant parts of A 's behavior are neatly explained by assuming that her explicit mental representations are governed by false identity claims. Since the theorist is not restricted to using the language of the subject under discussion she is free to use the identity sign to capture these facts in a clear and precise fashion. And nothing seems to compete, in terms of clarity, with using the identity sign in these cases.

1.3 Proper functions and malfunctions

Given that I've shown that the Frege model best explains the mental state of confusion, we don't need to construe confusion as "an error of its own kind" (Millikan 2000: 173), as opposed to a false belief. Even so, I'll now go on to argue that confusion in fact constitutes a distinctive kind of malfunction, one that we can even characterize in something like Millikan's own terms. Upon reflection, false beliefs aren't all created equal. To illustrate, if someone endorses *trivialism* and honestly believes *that everything is true* she makes, arguably, an error of a special kind that consists in having a specific set of false beliefs (cf. Priest 2006: ch. 3). Being a trivialist ought to have quite specific detrimental effects on one's ability to cope with a changing world. Thus the Fregean can say with Millikan that confusion has peculiar distorting effects on speakers' thought and talk.

In this section I explain, using Millikan's teleosemantic framework, in what sense confused identity is a cognitive malfunction. It is not my intention to endorse teleosemantic or biosemantic theories of intentionality. It seems, however, that Millikan's notion of 'proper function' can be adopted without accepting much else from her picture of language and communication (see, especially, Origgi & Sperber 2000).¹⁸ Subsequently, in §1.3.2, I propose an intuitive, Gricean definition the proper function of singular terms in linguistic communication and argue that confused identity has specific distorting effects on that function.

1.3.1 Proper functions introduced

What is a proper function? The notion is, of course, borrowed from evolutionary biology (see Millikan 1984, 1989b, 1989a; also Godfrey-Smith 1994; Papineau 1984, 1987). One item can serve many different functions at the same time but usually only a subset of these functions actually helps explain why the item continues to be reproduced. Consider the human heart. It seems to have many different functions. For example, the heart makes the human body

¹⁸More to the point, there is no need to accept the most controversial aspects of teleosemantics. The following discussion only assumes that there are *general* cognitive or communicative mechanism that have proper functions in the evolutionary sense, and this ought to be common ground between naturalistically inclined theorists. I do not assume, for example, that the particular *contents* of specific concepts or of token linguistic expressions can be explained solely in terms of their proper function (see, e.g., Pietroski 1992).

heavier, emits a low thumping sound, and pumps blood. Only the last is plausibly thought of as the proper function of human hearts, because it is historically responsible for the fact that hearts are reproduced. Importantly, an item that does not actually pump blood can still count as a heart. Malformed and malfunctioning hearts are still 'supposed to' pump blood, i.e. serve the proper function whereby their proliferation is evolutionarily and historically explained. Going further, Millikan applies the notion of proper function to biological and cultural items alike. Thus, according to her, linguistic devices (words, syntactic forms, etc.) have their own proper functions. For example, the proper function of the indicative mood is the production of a belief in the hearer that corresponds to the meaning of the indicative sentence uttered. The imperative mood has the proper function of producing compliant behavior. If I say "Pass the salt" and my interlocutor then passes the salt, my imperative utterance has successfully performed its proper function. The proper function can also be called the 'stabilizing' function of the item, since it is responsible for perpetuating the item in a community.

Origgi & Sperber (2000) argue in detail that Millikan's theory of linguistic devices is wrong because it is not necessary that imperative utterances typically or reliably produce compliant behavior, or even the hearer's desire to comply.¹⁹ On their view it is more plausible to say that imperative utterances function properly when the speaker succeeds in guiding the hearer towards the correct interpretation, so the hearer understands which course of action would satisfy the imperative utterance. Compliance is an additional step that depends entirely on the hearer's own beliefs, desires, and intentions. According to Origgi & Sperber: "Linguistic devices produce highly reliable responses, not at the level of the cognitive outputs of comprehension such as belief or desire formation, and even less at the level of behavioural outputs such as compliance, but at an intermediate level in the process of comprehension" (ibid., 161).

Although their own relevance-theoretic approach is plausible, Origgi & Sperber make an instructive mistake in interpreting Millikan's proposal. Thus their negative arguments are idle. Even if imperatives produced compliance only on very rare occasions, it might still be

¹⁹Cf. "On the view Millikan defends, comprehension *typically* consists in coming to directly believe what is asserted or in coming directly to want to comply with what is being requested" (ibid., 143). Also: "We follow Millikan in considering that the direct proper function of a linguistic device is what keeps speakers and hearers using and responding to the linguistic device in a *reliable* way, thus stabilising the device in a community" (ibid., 160). (Emphases mine.)

the case, at least on this theory, that their proper function is to produce compliant behavior. Millikan has emphasized that proper functions have nothing to do with typicality, reliability, or statistical averages (1989b: 21–22, 1989a: 93–94, 1984: 4, respectively). And this fits well with functional notions in biological theory. All that is required, on Millikan's view, is that an explanation of why imperatives are reproduced essentially invokes historical occasions on which imperatives in fact caused compliance. A sequence of such occasions would then explain why imperatives were 'selected for' by a process analogous to natural selection. Millikan defines 'normal conditions' as the actual conditions that have been historically needed for an item to perform its proper function. In the case at hand, the normal conditions might involve things like the hearer's understanding, the speaker's command of the hearer's language, and the hearer's willingness to do as told.²⁰ The fact that hearts function to pump blood is explained by a history of normal hearts pumping blood in the normal way under normal conditions. As Millikan (1984: 56) puts it: "If no token of the imperative mood ever effected more than an abortive attempt or intention to comply with it, it is clear that speakers would soon cease to use the imperative forms at all or to use them as they now do."

This does not establish that Millikan is right about the proper function of imperatives. But the *prima facie* plausibility of her account, even on this controversial point, should be recognized. Now, let us start applying the notion of proper function to identity confusions and, then, to the ways in which they can distort linguistic communication. On Millikan's account the distorting effects of confusion on basic cognitive processes are fairly straightforward. As mentioned above, she argues that a central task of cognition is to reidentify particulars and properties in thought and perception (see Millikan 2000 especially). Thus, on her view, we can postulate a basic cognitive mechanism in the mind/brain with the proper function of reidentification. The mechanism itself, let us suppose, has proliferated and is perpetuated in humans by a process of differential reproduction. The mechanism underlies many important cognitive tasks, such as recognition, expectation, and inference (cf. Lawlor 2001; Recanati 2012).

Perhaps this alleged mechanism is only postulated because of the observed effects of many more basic mechanisms having to do with memory and intelligence. Then so be it. But if we

²⁰Cf. Godfrey-Smith's (1994: 265–266) discussion of the neck-expanding display of the frill-necked lizard.

can describe such a mental mechanism, at some level of abstraction, combinatory and separatory confusions—as defined by the Frege model—are clearly the most characteristic ways in which it breaks down and fails to perform its proper function. The function is to keep track of particulars in the world and false identity beliefs give rise to systematic and characteristic errors. For example, if *A* expects Bill (or Biff, or Bill/Biff, or the referent of ‘Bill’ in *A*’s language, etc.) to knock on the door in five minutes while implicitly believing falsely *that Bill = Biff*, then *A*’s reidentificatory capacity is disrupted, simply because *A*’s belief makes the relevant identifications unreliable. Repeated and chronic confusion indicates that the underlying mechanism itself is impaired.

1.3.2 The proper function of singular terms in communication

Finally, let us apply the notion of proper function to singular terms in a natural language. Millikan holds that the stabilizing function of a *proper name*—a paradigm example of a singular term—is to “precipitate an act of identification of its referent” (1984: 80). This is intuitive enough, but I prefer to put it into its more explicitly Gricean garb and apply the definition to all singular terms.²¹ The following is a rough characterization to be fleshed out in much more detail in Chapter 4.²²

Singular terms

The proper function of a singular term *a* in some natural language is to provide the hearer with evidence for the speaker’s intention to refer to a particular object *o*, when the speaker utters *a* on a given occasion.²³

²¹Without Millikan’s approval, of course, but see Origgi & Sperber (ibid.) as well as Chapters 3 and 4 below. See also Bach (1987), Carston (2002), Grice (1989), Neale (1992, 2005, forthcoming), Schiffer (1972, 1987, 2003), Sperber & Wilson (1986/1995), Wilson & Sperber (2012).

²²See the definition of ‘Gricean speaker w-reference,’ p. 106.

²³More strongly, having this proper function is constitutive of being a singular term. But the present argument does not require this stronger assumption. Note, however, that there are substantive and interesting differences between the role or linguistic meanings of different types of singular terms. For example, indexicals and demonstratives seem to incorporate some type of contextual *perspective* on the referent: ‘you’ refers to the addressee, ‘that’ refers to an object the speaker takes to be salient, etc. Proper names seem to lack these perspectives (cf. Neale 2008; Rothschild & Segal 2009). These kinds of differences are not at issue here, as all singular terms share, at a higher level of abstraction, the function of evidencing referential intentions. And this is, I argue, their constitutive and proper function.

Intuitively, if all singular term tokens would become confused, speakers would soon cease to utter them or they would acquire some distinct function.²⁴ This intuition seems, however, to depend on the assumption that members of the linguistic community in question are all confused *in different ways about different things* or that some are confused and some are not. Matters are very different when a whole group of speakers, for example a whole scientific community, is uniformly confused about the identity of objects or properties to which they intend to refer in speech (Camp 2002: ch. 2; Evans 1982: ch. 11; Field 1973). As Hartry Field argues, using ‘mass’ as an example, such global confusion of two different properties or natural kinds makes the reference of a term metaphysically indeterminate. Although such cases are interesting in their own right, I want to leave them to one side. I want to argue, rather, that in cases where the confused agents are perfectly capable—by being a bit more careful or discerning—of seeing things aright, the evidential function of singular term utterances becomes ‘corrupt’ in principle. More precisely, the confused speaker’s referential intention is conflicting and the evidence provided by the utterance of the singular term is corrupted. And the corruption is not due to Fieldian indeterminacy.

Thus, to focus on ‘local’ rather than ‘global’ confusions, assume that the speaker’s (A) language is a c-language and the hearer’s (B) language is an unconfused t-language. And start by putting the argument in terms of combinatory confusion:

P1 A makes an utterance U_c of ‘...Bill_c...’ to B on a particular occasion.

P2 A believes that $Bill_t = Biff_t$.

P3 $Bill_t \neq Biff_t$.

P4 A intends U_c as evidence for A’s intention to refer to $Bill_t$.

P5 Given (P2), A also intends U_c as evidence for A’s intention to refer to $Biff_t$.

P6 Since B knows (P3), B only takes U_c as evidence for either (P4) or (P5).

C7 Therefore B will not understand U_c as it was intended by A.

²⁴This point is elaborated in §4.1.4 below.

Thus, the mere fact that *A* is confused but *B* is not—in the relevant way—disrupts the evidential proper function of *A*'s utterance of a singular term. The intention is necessarily conflicting and this corrupts the evidence. Each step of the argument, (P1) through (C7), is unassailable, it seems to me. Two doubts might be raised, however. First, how are we sure that (P5) is true? Well, just consider a normal unconfused analog. I know the truth about Cicero being Tully. If I utter '...Cicero...' intending to refer to Cicero, who is identical to Tully, my utterance is also evidence that I intend to refer to Tully. Since my belief is true, no problem seems to arise. By the same token, U_c is intended as evidence for *A*'s intention to refer to two distinct objects, in virtue of *A*'s false belief (P2).

Secondly, if the reader has doubts about (P6) it is easy to point out that it follows from an unobjectionable principle:

PR1 If *B* takes *E* as evidence that *A* intends to refer singularly to *a*, and *B* knows that $a \neq b$, then *B* will generally not take *E* as evidence that *A* intends to refer singularly to *b*.

But this principle suggests a question. What happens if *B* actually *knows* that *A* is confused, i.e. knows *both* (P2) and (P3) in the above argument? This situation also gives rise to disruption of a similar kind.

Assume (P1)–(P3) as before, but change the rest as follows:

P4* *B* knows both (P2) and (P3).

P5* U_c is evidence (for *B*) that *A* intends to refer Bill_{*t*}.

P6* Given that *B* knows (P2), U_c is evidence (for *B*) that *A* also intends to refer to Biff_{*t*}.

P7* Given that *B* knows (P3), U_c is *not* evidence (for *B*) that *A* also intends to refer to Biff_{*t*}.

C8* Therefore, U_c gives *B* conflicting evidence as to *A*'s intention to refer a particular object.

Again, someone might want to cast doubt on (P7*)—similar to (P6) above. But (P7*) also follows from the general principle formulated in PR1. In other words, if *E* is evidence that the object intended is *a* and *B* knows that $a \neq b$, then, generally, *E* is not evidence for *B* that *b* is also intended. If this is right, combinatory confusion as such disrupts the evidential proper

function of singular terms in linguistic communication. In this particular scenario, the hearer lacks a principled reason to determine the speaker's referential intention precisely. A rational hearer, in a minimal context in which nothing but the subutterance of the proper name is considered, has equal reason both to take the evidence as indicating one object (e.g. Biff_t) and as not indicating that very object—because it then indicates some distinct object instead (e.g. Bill_t). This is what it means to say that confusion creates conflicting evidence for referential intentions.

Should we say instead, perhaps with Millikan on our side, that U_c does in fact provide good evidence for the intended referent being the Bill_t/Biff_t amalgam? I think not. To state the point briefly, uttering an ordinary proper name of English, for example, would almost never constitute good evidence for an intention to refer to such a gerrymandered object. Unless, perhaps, both speaker and hearer are confused in the exact same manner (more on that below). If I alone have the crazy belief that the apple in my hand is really a mereological fusion, o , of the apple and the Empire State building, I cannot expect my audience to get the reference to o by simply uttering 'This apple tastes great!' And if it is true, as I think it is, that communicative intentions are constrained by doxastic states then, normally at least, I cannot form the intention to refer to o since I may believe, in such a context, that it is impossible for the audience to comprehend me as referring to that gerrymandered object.²⁵

Parallel arguments can be given in the case of separatory confusion. I hope that it is fairly clear how the arguments would go. Thus I skip the first type of argument and go straight to the second, where the hearer is assumed to be aware of the speaker's false identity belief.

P1 A makes an utterance U_c of '...Superman_c...' to B on a particular occasion.

²⁵The literature contains a plethora of weaker and stronger versions of this constraint. For the purposes of my argument I only need this relatively weak formulation, so I stick to it throughout. Grice (1971), Harman (1976), and Velleman (1989: ch. 4) maintain that one must believe that one will in fact V . Brand (1984) says that one must believe that it is possible for one to V —this seems to have been Aristotle's view (cf. *Nicomachean Ethics*, 1111^b25). Bratman (1984, 1987), Davidson (1985: 215), Donnellan (1968), Mele (1992, 2001), and Pears (1985) all seem to fall somewhere in between these two extremes. For example, Mele (1992: ch. 8) argues that one normally cannot intend to V unless one lacks the belief that one will *probably* not V . Two dissenters, Anscombe (1957: §52) and Thalberg (1972) claim that one can actually intend to V while believing that one will not in fact V .

- P2 A believes that $\text{Superman}_t \neq \text{Clark Kent}_t$.
- P3 $\text{Superman}_t = \text{Clark Kent}_t$.
- P4 B knows both (P2) and (P3).
- P5 U_t is evidence (for B) that A intends to refer to Superman_t .
- P6 Given that B knows (P2), U_t is evidence that A does *not* intend to refer to Clark Kent_t .
- P7 Given that B knows (P3), U_t is evidence that A does intend to refer to Clark Kent_t .
- C8 Therefore, because of A 's confusion in (P2), U_t gives B conflicting evidence as to A 's referential intention.

Again, a critic might want to target premise (P7). But as before, (P7) follows from a sound general principle: if E is evidence that A intends to refer to a , and $a = b$, then E is evidence that A intends to refer to b . Bear in mind that the notion of *intending to refer to a* appears to be a transparent one. The principle would *not* hold water if it were substituted for an opaque notion, such as *intending to refer to a as a* or *intending to refer to a as b*. Similarly, while *seeing* is transparent, *seeing as* is opaque. If I see John and John is my cousin, I also see my cousin. But seeing John as John does not imply seeing John as my cousin. By the same token, if I intend to refer to John and John is my cousin, I also intend to refer to my cousin. This remains true even if I utter (confused) sentences like 'My cousin and John are not the same person.' But my false beliefs may, as in the two cases above, create a conflict in my referential intention and corrupt the evidential proper function of the singular term.

Someone like Field may object that my whole argument has a hidden false assumption, namely, that the locally confused speaker can have a determinate referential intention at all. Local confusion is just like global confusion: it gives rise to metaphysical indeterminacy. Field, I imagine, would go on to argue that there is no point in talking about *conflicting evidence* for an *indeterminate* conclusion. On this account, then, the combinatorily confused speaker, for example, has an intention that refers indeterminately to Bill_t , Biff_t , and the $\text{Bill}_t/\text{Biff}_t$ amalgam, and the utterance U_c is good evidence for this indeterminate intention.

The objection misses the point of the present exercise. If we consider ordinary proper names and perceptual demonstratives, rather than natural kind predicates like ‘mass’ in Newton’s language, it is clear that even the confused speaker *can*, on occasion, have a very determinate intention to refer to an object about which he is confused. But, in those cases, the part of the linguistic evidence that consists in an utterance of a confused token of a singular term will still be corrupt, just with a diminished practical upshot. Here is the kind of case I have in mind. A confuses $Bill_t$ with $Biff_t$ but stands right in front of $Bill_t$ and utters ‘ $Bill_c$ is right there’ while pointing directly at him. Assume also that A hasn’t seen or thought of $Biff_t$ for quite some time; $Biff_t$ ’s contextual salience is below zero. In this case, I suppose, A definitely intends to refer to *that man there*, i.e. $Bill_t$. Without doubt, A also intends to refer to $Biff_t$, but this is such an insignificant part of A’s communicative intention that it seems irrelevant. So the evidence provided by the singular term token is, strictly speaking, conflicting. Practically speaking, the conflict is quite inconsequential because of the context.²⁶

To summarize, the mere presence of identity confusion with respect to object *o* will make it such that a singular term token, that is intended by the speaker to refer to *o*, is unable to perform its stabilizing proper function. It is reasonable to suppose, then, that the reproduction of such terms in natural language needs to be explained by invoking a sequence of utterances where such confusion was not present at all. And if the notion of a proper function is a good instrument to use in defining the encoded or literal meanings of natural language expressions, this conclusion will arguably have vast implications for the way in which one ought to theorize about reference and language more generally.

Conclusion

I have argued that identity confusions can be defined as a mental state, characterized by certain propositional attitudes. Minimally, the confused agent must either not believe a true proposition about identity or believe a false proposition about identity. The most serious

²⁶I intend this to be a case where nothing fanciful or abnormal is going on—one of a class of paradigm cases of referring with a proper name—and this makes it different from the atypical cases to be discussed in detail in the next Chapter. I point this out to make clear that I am committed to the view that our ‘intuition’ about a case like this *can* indeed support—without conclusively establishing—one theory rather than another.

objections to this account can be staved off by adding that the beliefs can be implicit, but still, the agent must stand in some explicit attitude relations to the object(s) in question. Furthermore, I showed how this mental state has disrupting effects on the proper function of singular terms in communication. If I am justified in thinking that confusion corrupts the function of referring expressions, many influential arguments in philosophical semantics, based on the coherence of various ‘semantic puzzles,’ can be substantially undermined. This is the topic of the next Chapter.

Chapter 2

Puzzle-driven semantics

Introduction

How have theorists traditionally conceived of the basic theoretical task of the philosophy of language? I argue that the dominant conception since Frege has been, implicitly but often quite explicitly, that there are semantic puzzles in need of philosophical solutions. This is particularly clear in theories of singular terms. I call the approach ‘puzzle-driven semantics’ and argue that it has been fruitless, resting on flawed methodology.

First, on the basis of the conclusions from Chapter 1, I show how this tradition can be defined and demarcated by its core commitments. I show that the research program thus defined ought to be abandoned or revision in fundamental ways. Furthermore, I argue in a general manner that the philosophy of language should be ‘explanation-driven.’ But only in Chapter 3 do I start developing my own view on how the basic explanatory task, identified here, ought to be fulfilled. In the present Chapter, I argue that this latter approach has a well-defined and worthwhile goal, to which solving semantic puzzles has no direct application. Surely there is by now a vast jungle of different approaches to basic semantic questions and, moreover, it seems like many of them can lay claim to being worth one’s trouble.¹ But for concreteness I focus only on two frameworks of this sort.

In the the main part of the Chapter (§2.2) I focus on three puzzle-driven arguments, some of which have been hugely influential—e.g. Kripke’s argument for distinguishing semantic and speaker reference. Each argument aims to draw major semantic conclusions from puzzles involving confused speakers, but, as I show based on the prior discussion, the inference can

¹See, especially, Devitt (1996: ch. 2) for an opinionated survey.

be defeated in each particular case. For comparison I also discuss an argument that seems to share many characteristics of a puzzle-driven methodology but the argument still goes through (§2.3). Interestingly, this is because the relevant speaker does not really satisfy my definition of confused identity from Chapter 1 but this definition is crucial in characterizing the puzzle-driven approach. Finally, in §2.4, I evaluate a puzzle-driven argument in the theory of *anaphora*, due to Scott Soames, reaching the same conclusions as before.

2.1 What is puzzle-driven semantics?

Philosophy of language has long been dominated by a distinctive set of methods, goals, and core assumptions. Recently, for example, it has become popular to characterize philosophy of language in terms of the so-called *method of cases*. On this method a theory of some phenomenon, say the meaning of expressions of type *T*, is considered true if it is best supported by the intuitions that competent users of *T* have about the meaning of *T*-expressions across actual and possible cases (Mallon *et al.* 2009: 338). But many have argued that philosopher's reliance on intuitions about such cases is unjustified and unjustifiable. If intuitions are mere snap-judgments prompted by, e.g., Kripke's examples in *Naming and necessity*, it is quite unclear why they should be taken seriously as data. Such intuitions are nothing if not fallible. This observation has lead theorists along at least two distinct routes. One is the road to experimental philosophy, where theorists pool together the intuitions of the folk and, where these are relatively robust and uniform, continue to use them in support of philosophical theories (e.g. Alexander 2012). But one can also deny the assumption that semantic theories are supported (or defeated) by mere appeals to intuition (e.g. Cappelen 2012; Ostertag 2013).

The truth of is more complicated. Everyone, also experimentalists, appeals to individual intuitions at one point or another. As Devitt (*ibid.*, 73) argues, intuitions are needed to get the ball rolling: expert intuitions help to identify the subject matter of semantics.² But these may

²A note on terminology. I use '(philosophical) semantics' and 'metasemantics' in a broad way to indicate the study of meaning in natural language, especially the metaphysics of meaning. I accept the gist of Burgess & Sherman's (2014) three-fold distinction here. First, *basic metasemantics* asks in virtue of what expressions have semantic properties. Secondly, *the theory of meaning* is concerned with discovering the *nature* of the meaning relation: what constitutes the fact that an expression has a given semantic property? (Note that this distinction

be revised and refined as more progress is made. Thus I do not think that the *problem* with the method of cases has anything to do with its appeal to intuitions as such. The problem has more to do with the underlying assumption that the philosopher's job is to systematize and explain intuitions about particular types of cases (Devitt 2012a). Now, according to explanation-driven semantics, described in detail below, this project has some legitimacy, but only for cases where the mechanism under scrutiny is performing its normal function in its proper setting.³ In my view, an unhealthy concern with cases involving confused agents, and with intuitions about such cases, has been an impediment to progress in the philosophy of language.

Armed with the notion of identity confusion as it was defined and defended in Chapter 1, I propose to define puzzle-driven semantics as follows.

Puzzle-driven semantics

A semantic theory is puzzle-driven if its basic task is to provide a coherent system with which to describe the relevant semantic features of cases involving utterances of speakers who suffer from confused identity.⁴

Let me clarify. First, even if providing such a descriptive system is considered a 'basic semantic task' it is, of course, not necessarily thought of as the only possible task. Theorists in this tradition assume, often explicitly, that the resulting theory applies directly to other types of cases—specifically, cases where the speaker does *not* have false identity beliefs. This, as we shall soon see, is doubtful. Second, the semantic features in question can be of many different kinds: reference, truth, meaning, content. Naturally, our concern will be with what a speaker refers to in producing utterances containing confused singular terms.

assumes a difference between natures and grounds (cf. Audi 2012: §3.) Thirdly, *the metaphysics of semantic values* aims to discover the nature of the semantic properties themselves. If they are propositions, for example, are they structured/unstructured, Fregean/Russellian? But my project belongs, at different points, in all three categories. And, of course, the burgeoning science of linguistic semantics provides us with clues about how to identify the 'meanings' in the first place.

³Similarly, Peter Hacker (2007: 309–310) argues that intuitions are particularly useless when considering farfetched and nomologically impossible cases—such as consulting intuitions about brain-transplants to reach conclusions about personal identity—since one really has no idea what to say about such fanciful scenarios. See also Hanfling (2000: ch. 4).

⁴For some examples, see, Ackerman (1979); Bealer (1993); Evans (1982); Fiengo & May (2002, 2006); Fine (2007); Kripke (1979, 1980); Reimer (1991a, 1992b); Sainsbury (2002); Salmon (1986a).

Furthermore, theories of this sort will appear to have many traditional theoretical virtues. First, if theory T captures the semantics of more confusion cases than theory T' —i.e. T solves more puzzles—then it seems to win out on descriptive adequacy. Second, if a theory is ontologically conservative it is better than another that introduces new types of entities—such as Fregean senses, as the case may be. Third, theories can provide a more or less compelling explanations of the data; that data being judgments about cases involving confused speakers. But, as I will argue, such apparent theoretical features must be grounded in a viable conception of the explanatory goals of philosophical semantics.

As indicated, the goal of puzzle-driven semantics is often described as that of giving solutions to philosophical or semantic puzzles. Kit Fine (2007, 2010b) is helpfully explicit on this point. His ‘semantic relationism’ is ultimately a *solution* to “Russell’s antinomy of the variable, ... Frege’s puzzle in its various guises, and ... Kripke’s puzzle about belief” (2007: 5).⁵ But what does it mean to ‘solve’ such ‘puzzles’ and why exactly should we try to do it? In other words, assuming that the definition of puzzle-driven semantics is correct, why should we try to discover a catch-all description of all conceivable confusion-type cases?

There is no good reason. Even if we had a successful theory of this sort the point of it would still be unclear. How can this be? For two reasons. First, it should be obvious that there are indefinitely many possible systems of description that can cover all of the conceivable confusion-type cases in one way or another. Since Frege, many such have been proposed and some developed in painstaking detail. Fine’s semantic relationism is a telling example. But none of these solutions have garnered widespread support and thus, it seems, more is needed for their adjudication. So, solving the puzzles cannot be the most basic task in the metasemantics of singular reference. To earn their keep, such solutions ought to help us with other questions that are, arguably, more basic. However, as I hope will become clear, it is quite implausible to think that they provide any aid in tackling other more basic questions in the philosophy of language.

Secondly, it seems uncontroversial that the task of theorists working in semantics and prag-

⁵It is run-of-the-mill, standard practice in philosophy of mind and language to argue that the way a theory takes care of Fregean puzzles is very strong evidence in its favor. For a good recent example, see Hanks (2011).

matics has something to do with *explaining* linguistic behavior. In particular, one is interested to know how it is possible for people to utter—by speaking or writing—linguistic expressions and thereby succeed in expressing and communicating their thoughts and ideas to others.⁶ And the emphasis here is on expressive or communicative *success*: accounting for the mechanism of linguistic communication in cases where everything goes *as planned*. Otherwise we mistake merely accidental forms of success for the real thing, and the theory suffers accordingly. One of the motivating ideas of this dissertation is that there is more to actual communicative success than merely having a speaker who utters something on some occasion and a hearer who then recognizes what the speaker intended to communicate. More specifically, we need to formulate specific cognitive constraints on the proper performance of the speech act of referring with a singular term (more on this in Chapters 3 and 4 below).

In the last Chapter, I argued that utterances containing ‘confused’ singular terms are abnormal or corrupt in a specific sense. If this is accepted a special justification is called for if an account that is purposively designed to capture confused utterances is to be applied unaltered to normal, ‘core’ cases involving only unconfused utterances of singular terms. For it is only in the latter case that the mechanism of communication is working properly and normally.

A rough analogy will help here. Designing one’s theory so that it fits the semantic features of as many confusion cases as possible is like developing a psychological theory of face perception that applies only to prosopagnosics and, then, imagining that it must *also* hold true of normal subjects.⁷ Of course, I’m not suggesting that confused identity is a neurological disorder akin to acquired or congenital prosopagnosia. There is no such thing as congenitally confused identity. But this is not because the specification of confusion depends on objects external to the mind, for prosopagnosia is external in a similar way: face blindness is not a disorder in a world without faces.⁸ More to the point, a theory of face perception is a theory

⁶Neale (2004: 71–72, forthcoming) calls this the *Master question* in philosophy of language: “How rich of an explanation can we provide of our capacity to express and sharpen our thoughts, and to communicate information about the world and about our beliefs, desires, plans, commitments, hopes, fears, and feelings so efficiently—so quickly, systematically and consistently—using various noises, marks, and gestures?”

⁷People suffering from prosopagnosia can, often with considerable difficulty, learn to recognize faces almost as quickly as others. But the process is usually quite artificial, involving the conscious use of contextual clues and various heuristics. See, e.g., Sacks (2010: 90–91).

⁸Yet it should be noted that prosopagnosics often have other problems as well, e.g. with

of the perceptual processing of faces in normal subjects (i.e. subjects not suffering from any relevant neurologic, psychiatric, or neurophysiological disorder). Abnormal cases contribute to this theory, but only indirectly, as they are not part of what is primarily to be explained. My suggestion is that confusion should be treated similarly in semantic theorizing. It is basically a pragmatic performance error in the Chomskyan sense—no matter how common or intrinsically interesting it turns out to be.⁹

One worry needs to be addressed before going any further. It may seem like my characterization of this alleged tradition is unfair. It may seem like there are many ways to pose something akin to Frege's puzzle without invoking any agents who satisfy the so-called Frege model of confused identity. The impression is wrong. To see this, let us look at a trimmed down version of Fine's (2007: 34–35) basic Fregean puzzle. He does not mention any speakers or hearers in stating the puzzle itself, or so it may appear. Basically, he says that there seems to be a semantic difference between (7) and (8).

(7) Cicero is Tully,

(8) Cicero is Cicero

Adherents of direct reference think that if there is any semantic difference between the names 'Cicero' and 'Tully' there is also a referential difference. But there clearly is no referential difference. Solving the puzzle would be to describe the difference or apparent difference between (7) and (8). Traditionally, Fregeans have posited a level of meaning distinct from reference and referentialists have denied that there really is any semantic difference. Fine's own idea is to be a referentialist while also introducing coordinating links between distinct occurrences of coreferring singular terms.

But this is not really enough to motivate the puzzle. We also need to know *why* there seemed to be a semantic difference in the first place. Fine provides the missing link by saying that (7) and (8) "can convey different information to someone who understands both sentences" (2007: 34).

object-recognition and large-scale navigation, cf. Duchaine (2011).

⁹Robert Stainton (forthcoming) makes a very similar point but uses a different example. He points out that some patients who have had Broca's area surgically removed may still retain their linguistic competence. The brain can recruit new areas to serve the same functions as Broca's area did before. Despite this possibility, as Stainton, points out, Broca's area forms part of the normal explanation of linguistic competence.

But this fundamental difference—fundamental because the puzzle cannot be stated without it—may appear to have nothing to do with confused identity. It seems like *A*'s utterance of (7) can be informative to *B*, in a particular context, even if *B* is not confused with respect to the referent *o* of 'Cicero'/'Tully'. Just suppose that *B* has never had any propositional attitude the object of which is a proposition *p* expressed by uttering a sentence containing the name 'Tully' in virtue of which *p* has an *o*-dependent truth condition. But (8), assuming *B*'s familiarity with 'Cicero,' would *not* be informative in the same context.

On reflection, however, it is clear that this does not provide the motivation needed. An utterance of (7) would be an odd one in the context as described. It falsely presumes the hearer's familiarity with a practice of using the name 'Tully' to refer to *o* (where *o* = *Cicero*). And the most reasonable interpretation *B* can give is that *A* meant to utter something like: "Cicero is also called 'Tully'." If this is right there is no puzzle about the difference in informativeness: one sentence is uttered to introduce a new name, the other to assert a trivial identity. The puzzle only comes up if the presumption of the hearer's familiarity with the name 'Tully' is true.¹⁰ But then the Frege model of confusion must indeed be true of the hearer, i.e. the semantic difference is only puzzling if the hearer is separately confused with respect to Tully/Cicero. This remains so even if the hearer is *indifferent* as to whether 'Cicero' and 'Tully' refers to one or two individuals. For, in that case, the hearer lacks the belief that Cicero is Tully, and lacks the belief that Cicero is not Tully. And the former is sufficient for the Frege model to apply.

It should be noted, here, that in Fine's way—and Frege's, for that matter—of posing the puzzle, the speaker is only a theatrical device. The puzzle arises for the hearer considered in isolation, given her confused mental state. It should be posed, like we did with Lois Lane and Superman in the last chapter, as a puzzle about the contents of the hearer's potential thoughts or utterances involving the relevant singular term. Thus it is no objection to my characterization of puzzle-driven semantics that theorists have only assumed that there are confused or ignorant *hearers* or *interpreters*, and that the *speakers* may be as ideally knowledgeable as one could wish. Hearers are usually speakers as well.

My alternative to puzzle-driven semantics is explanation-driven semantics. Such a semantic

¹⁰Salmon (1986a: 60) also makes this point.

theory, as noted above, is ultimately driven by a desire to explain the fact of successful linguistic communication.¹¹ As many theorists in the Gricean tradition in semantics and pragmatics have argued it is important, in order to carry out this kind of explanatory project, to be clear about which semantic categories are taken as metaphysically basic. We need to know which semantic notion is plausibly thought of as explanatorily prior to any other that we might care to posit while constructing the theory. Now, there are three relevant notions that one finds in the literature. First, there is ‘speaker meaning,’ or the meaning intended by the speaker on a specific occasion of utterance. Second, there is privileged interpretation, or the meaning that an ideally rational hearer, under normal or ideal conditions, would arrive at in interpreting some utterance on a given occasion. And finally, there is the conventional or encoded meaning of expression types in particular natural languages. Call this ‘linguistic’ meaning. Let us consider each notion in turn.

Speaker meaning: Intention-based semantics in the tradition of Grice (1989) is wedded to the second option.¹² Intention-based theorists claim, plausibly in my view, that the basic explanatory notion is that of speaker meaning, which is constituted by certain specific communicative intentions. The specific nature of said intentions is, of course, a matter of dispute. But the following principle, which will simply be called ‘intentionalism,’ is more or less common to them all.

Intentionalism

If speaker *S* says and means that *p* in uttering sentence σ in language *L* on some occasion, *p* must be (i) constitutively determined by *S*’s communicative intention in uttering σ on that occasion and, (ii) *p* must be compatible with the abstract linguistic meaning of σ in *L*.

¹¹Even if one endorses an explanation-driven approach it doesn’t follow that philosophical puzzles, or confused agents, are completely irrelevant to semantic theory. Just like performance errors can constitute evidence for features of an underlying linguistic competence (Chomsky 1980: 200–201), taking the possibility of confusion into account can lead to insights and evidence of different sorts. Perhaps Russell’s (1905: 484–485) famous remark about the importance of stocking the mind with as many puzzles as possible can be understood in roughly this way.

¹²See also Bach (1987); Bach & Harnish (1979); Bennett (1990); Carston (2002); Donnellan (2012); Harris (2014); Loar (1986); Neale (1992, 2005); Schiffer (1972, 1987, 2003); Sperber & Wilson (1986/1995); Wilson & Sperber (2012); Strawson (1971). Some theorists disagree with the following, even as an interpretation of Grice. See, e.g., Saul (2002).

What ‘compatibility’ amounts to here is, of course, a vexed issue that I won’t go into until §5.1 below. Notice also that, on the Gricean view, (ii) is dropped in explicating other parts of what the speaker *means*. This latter notion incorporates the “total signification” (Grice 1978: 41) of a linguistic utterance, including what is merely implicated and, thus, not linguistically constrained.

The upshot, for our purposes, is as follows. If *S* actually intends to say (and mean) *that p* by uttering σ on some occasion, and *p* is σ -compatible, then *p* is the content of what *S* said on that occasion. Even if, based on all the available evidence, the hearer *H* simply cannot understand *S* as saying *that p*, still, the content of what *S* said is *p*. This is the content that *S* intends for *H* to understand and on a very compelling picture of successful communication *H*’s job is to figure out *S*’s communicative intention on the occasion of utterance. *H*’s job is not to figure out what σ means in the language, although knowing this would usually help quite a bit. Further, even if an ideally rational agent would understand *S* as saying *that q* on this occasion, what *S* said is still *p*. Importantly, it is part of the intention-based project in the philosophy of language that the meaning of σ in the language can only place constraints on the type of communicative and referential intention *S* can form in uttering σ .¹³ If linguistic meaning can be explained in terms of a prior notion of speaker meaning, which seems plausible, the latter stands a good chance of being explanatorily basic.¹⁴

Ideal meaning: Many theorists seem to go along with something like explanation-driven semantics without thereby endorsing intentionalism. In fact, it seems common to assume instead that an ideally reasonable, or ‘contextually determined,’ interpretation is the most basic determinant of content.¹⁵ For lack of a better name, let us call this view ‘interpretive contextualism.’

Interpretive contextualism

The content of what a speaker *S* says and means by uttering a sentence σ in language *L* on some occasion is constituted by the interpretation that an ideally rational and *L*-competent hearer would assign to the utterance on that occasion.

¹³Cf. note 25 in Chapter 1.

¹⁴For arguments in favor of intentionalism, see Bach (2005); Neale (2005: 175–204).

¹⁵See Reimer (1992b, 2004); Travis (1989: ch. 2, 1991); Wettstein (1984: 80–81).

Like intentionalism, it is plausible to think that contextualism usually determines a unique content for specific utterances. At the very least, different assumptions about the epistemic state of the ideally rational hearer will lead to different specific contents. Each content, however, will definitely be a full-fledged proposition with a truth condition. Clearly, the two views are in strict opposition to one another. One major motivation behind interpretive contextualism, it seems to me, is the desire to accommodate cases where speakers appear to *say* something without *meaning* it in any way whatsoever. The contextualist takes this possibility seriously, while the intentionalist, as here construed at least, is committed to rejecting it.

For illustration, consider this example from Perry (1997, 2012: 70).¹⁶ When Rip Van Winkle falls asleep on July 3, 1766, and doesn't wake up until twenty years later, he may utter

(9) I fell asleep yesterday,

intending to refer to July 2, 1766. But, according to Perry, he will *in fact have designated* July 2, 1786. This is because, on Perry's view, 'yesterday' in (9) simply refers to the day before the utterance occurs, no matter what (else) the speaker intends. The reference of the indexical is 'automatic.' An interpretive contextualist might explain this by pointing out that the 'automatic' reference is what an ideally rational interpreter would come up with in the context of Rip's utterance. Adherents of intentionalism will have to disagree and argue that the ideal interpreter would be making a mistake. I turn to this example again in §2.2.3 below.

Linguistic meaning: The sentence 'She's ready,' for example, has some encoded or compositional linguistic meaning as a grammatical clause in English. In this case the immediate constituents of the clause are [_{NP} she] and [_{VP} 's ready], the first, according to a popular theory, standing for a function from contexts to objects and the second for a function from objects to truth-values. On this picture of semantic composition, the sentence as a whole stands for a truth-value. Classically, theorists who believe this notion is explanatorily basic are drawn towards a third view of meaning determination that might be called 'objective contextualism.' The more traditional term would be *indexical semantics* (cf. Kaplan 1989b; Lewis 1980; Rothschild & Segal 2009).¹⁷ On this view, it is simply a semantic fact about a token indexical, say,

¹⁶Based on a short story by Washington Irving published in 1819.

¹⁷There is much variation among theorists here, of course. Many suggest that the 'external'

that it refers to a given individual in the context of utterance—the context being a set of objects, called an ‘index,’ such as the speaker, a time, and a place. So, in principle, both the speaker and the ideally rational hearer could be wrong about what is actually referred to in context.

But, as many have argued, the category of the meaning of sentence σ in-the-language is so abstract and minimally informative that it underdetermines the intended interpretation of any or most σ -utterances (cf. Unnsteinsson 2014).¹⁸ I turn again to this as a substantive issue Chapter 3 below. For now, I assume that linguistically encoded content is too minimal to be explanatorily basic in communication.

2.2 Three shoddy arguments?

In this section I argue against three influential arguments for substantive semantic theses. Crucially, the source of the problem is always the theorist’s commitment to a puzzle-driven approach. This is strong evidence for the view that the approach is, on the whole, misguided. To make my own arguments as general as possible, they are neutral between intentionalism, interpretive contextualism, and objective contextualism. All three, arguably, give rise to the very same problems. The discussion of Marga Reimer’s argument in §2.2.2 is a necessary exception; in that case my argument clearly supports intentionalism as against the alternatives.

To keep the discussion coherent, however, I describe the examples in terms of intentionalism, explaining the application to interpretive contextualism only in passing. This unmasks my loyalty but it is fairly easy, I hope, to see how the same argument applies with different assumptions. Applying the argument to objective contextualism would be, I submit, entirely or ‘objective’ context—thereby excluding the speaker’s communicative intention—is constitutive of what the speaker says and means (or specific aspects thereof). See Gauker (1997, 2003); Perry (2012); Recanati (2004); Stanley (2007). Others emphasize linguistic conventions (e.g. Devitt 2013a; Lepore & Stone 2015). Many philosophers also combine different aspects of the three views stated here (cf. King 2013, 2014; Soames 2005, 2009a, 2010a: 173).

¹⁸See also Bach (1987, 1994a, 1994b, 2005); Bezuidenhout (2002); Capellen & Lepore (2005); Carston (1988, 2002); Chomsky (1977); Downing (1977); Grice (1989); Kripke (1977); Ludlow (2014); Neale (1990, 1992, 2004, 2005, 2007, forthcoming); Perry (1979, 1986, 2000, 2012); Predelli (2005); Recanati (1989, 1993, 2001, 2002, 2004, 2010); Reimer (2002); Schiffer (1977, 1978, 1981, 1992); Searle (1969, 1975, 1978, 1980); Soames (2009b); Sperber & Wilson (1986/1995, 1998, 2005); Stanley (2007); Travis (1975, 1985, 1989, 1997, 2008); Weiskopf (2007); Wettstein (1984); Wilson (2005); Wilson & Sperber (2002b, 2012).

analogous, but adding it explicitly to the mix would be far too cumbersome. At any rate, such a theory brings its own specific complications to be addressed in Chapter 3.

2.2.1 Kripke's theory of semantic reference¹⁹

Saul Kripke (1977, 2013: ch. 5) produced some quite influential arguments for a distinction between semantic reference and speaker reference. The distinction is in terms of speaker's idiolects. According to Kripke, certain conventions of the idiolect—given various facts about the world—determine the *semantic* referent of a designator in the idiolect.

His chief argument—which crucially involves a confused speaker—is as follows. Imagine a context in which *A* and *B* see a man in the distance raking leaves. The man happens to be their mutual acquaintance, Smith, but *A* and *B* both mistake Smith for another friend of theirs, called 'Jones' in their respective idiolects. *A* asks, 'What's Jones doing?' and *B* replies,²⁰

(10) Jones is raking the leaves

For *B*'s utterance of 'Jones' in (10) to satisfy the Frege model of confusion we need to be explicit that *B* simultaneously believes the proposition *B* would express by uttering, 'That's Jones' while pointing to the man in the distance. This is clearly implicit in Kripke's own story. And then it's true, on this occasion, that $\llbracket \text{That} \rrbracket \neq \llbracket \text{Jones} \rrbracket$ and that *B* believes that $\llbracket \text{That} \rrbracket = \llbracket \text{Jones} \rrbracket$. *B* is therefore confused about the identity of the intended referent. Now, Kripke tells us that "in some sense" (1977: 111, 2013: 118) both *A* and *B* have referred to Smith and *B* has said something true about Smith, namely that he is raking the leaves. Let us accept this intuitive judgment for the time being and call it 'Kripke's assumption.' But how do we account for its truth? This is where speaker reference and semantic reference occupy centre stage.

Kripke's argument has the form of an inference to the best explanation. He is ultimately concerned with arguing against theories that postulate ambiguities in definite descriptions

¹⁹Many thanks to Ben Phillips and Eliot Michaelson for pressing me to clarify the argument in this section.

²⁰I changed the example so as to set aside issues about implicit reference and aphonic expressions, cf. Neale (forthcoming). This is actually how Kripke formulates the example in his Locke lectures (2013: 118), which is different from his formulation in "Speaker's reference and semantic reference," (1977: 111).

(Donnellan 1966). So he starts by dismissing the implausible idea that his assumption can be accounted for by postulating an ambiguity in the name ‘Jones.’ Of course no one would say that the above example is evidence that in *B*’s idiolect the name ‘Jones’ is sometimes the name of Smith and sometimes the name of Jones. Kripke proposes a better explanation of his own assumption by making a distinction:

Semantic reference

In an idiolect L_A , spoken by *A*, the semantic referent of a designator d is the object o such that *A* has a *general* intention to refer to o whenever *A* utters d in L_A .

Speaker reference

On a given occasion of utterance the speaker referent of a designator d in L_A is the object o such that *A* believes, rightly or wrongly, that o is the semantic referent of d and *A* has a *specific* intention to refer to o by uttering d on that occasion.

With this distinction in place, Kripke claims that in uttering ‘Jones’ in (10) *B* must have intended, on that occasion, to refer to the semantic referent of ‘Jones.’ But *B*, as a matter of fact, didn’t refer to the semantic referent of ‘Jones’ and referred to the semantic referent of ‘Smith’ instead. This explains, according to Kripke, the sense in which *B* spoke the truth about Smith. The speaker can refer to o by uttering d even if they wrongly believe that o is the semantic referent of d . And of course we can readily explain the fact that there is *also* a sense in which *B* didn’t refer to Smith: the (semantic) referent of the name ‘Jones’ in *B*’s idiolect is not Smith, it is of course Jones because *B* presumably has the general intention to refer to Jones by uttering ‘Jones.’ As Kripke puts it, usually speaker reference and semantic reference converge but there are cases in which they come apart.

First I will register my agreement that, if this is all true, then it does indeed explain Kripke’s assumption. But there is a better explanation and the example doesn’t compel one to accept the notion that designators—even if relativized to idiolects—stand in semantic reference-relations to objects. Kripke mentions only one alternative explanation, which is particularly implausible, i.e. that a proper name like ‘Jones’ is ambiguous, sometimes referring to Jones and—when the

speaker is confused—sometimes to Smith.²¹ This makes his case appear stronger than it is. The following alternative is much simpler and more plausible than Kripke's own hypothesis.

By uttering (10), let us assume, *B* intends to communicate the proposition *that o is F* to *A*, *o* being the individual that they both believe they can see in the distance (and each believes that the other can see him as well).²² Of course, $o = \textit{Smith}$, but they don't know this. *B* believes truly that *A* is disposed to understand *B* as intending to refer to *o* by uttering 'Jones' on this occasion. Moreover, *A* believes that *B* believes that *A* is disposed to understand *B*'s utterance of 'Jones' in this context as intended to refer to *o*. Thus *B* and *A* mutually believe that, in this context, an utterance of 'Jones' is intended to refer to *o* and their beliefs are true. This suffices to explain Kripke's assumption that *B* refers to Smith, since $o = \textit{Smith}$, without appealing to a notion of semantic reference. But remember that even Kripke's notion of speaker reference incorporates the notion of semantic reference.

This is, of course, not enough. It should be objected that I have only shifted the burden of explanation to another issue. How do we explain the *other* sense in which *B* refers to *Jones*, and not to *o*, by uttering (10)? As Kripke might put it, the words themselves seem to refer, on this occasion, to Jones, who is nowhere to be seen.

The best way to explain this, I think, is by first constructing a related context for comparison. Call the original context '*C*₂' and imagine an earlier interchange, *C*₁. Here, *A* and *B* are walking just a few moments before they see a man raking leaves in the distance. But they anticipate seeing Jones, their mutual acquaintance. *A* asks, 'What do you think Jones will be doing when we see him later?' *B* replies,

(11) Jones will be raking the leaves

I take it that *B* has obviously referred to *Jones* here and not to Smith, even if Smith happens to be the individual referred to later in context *C*₂. What exactly explains the difference between *C*₁ and *C*₂?

²¹It should be noted that according to many theorists names are ambiguous, but in a very different way; name *n* is then ambiguous only in the sense that there are many individuals who truly bear *n*. So an unambiguous name is just a name (type) borne by only a single person. See §3.2.2 below.

²²In Chapter 3 below, I try to provide this assumption, and others of its ilk, with a theoretical basis.

If intentionalism is accepted, and linguistic comprehension is basically intention recognition, the answer is simple. In C_1 the most important evidence provided by the speaker as to their referential intention is the fact that they uttered the proper name ‘Jones.’ And in C_1 it is mutually believed by the participants that ‘Jones’ is intended and understood as referring to Jones, their mutual acquaintance. In C_2 , however, there are *two* pieces of evidence provided by the speaker when they utter (10). Both are extremely salient and important. The first is that B uses the name ‘Jones.’ The second is that B indicates clearly that the individual seen in the distance is the one to which B intends to refer by uttering ‘Jones.’ Since it is assumed that Smith is the one raking the leaves in both contexts it is clear that the evidence provided by B in C_2 is conflicting. B isn’t aware that the individual o is Smith and believes o is really Jones. And thus, part of the evidence suggests that B has a *specific* intention to refer to Jones and part of the evidence suggests that B has a *specific* intention to refer to o , who happens to be Smith. This fully explains Kripke’s assumption—without mentioning semantic reference at all—because the evidence is most naturally construed as in conflict about the speaker’s *specific* referential intention and not any kind of *general* intention. And this specific intention is internally conflicting because of the speaker’s confused mental state.²³

Briefly, Kripke’s own distinction displaces the real point of conflict in the speaker’s referential intention, and a different explanation is, then, needed. Speaker always aim primarily to convey their specific referential intentions, not any general referential intention—though they definitely have such intentions. Correlatively, the hearer’s goal is to interpret the speaker’s specific intention on the occasion of utterance. The immediate goal, adopted to achieve success in communication, is not to discover any general intention—i.e. semantic reference—but, again, such information is quite useful when available. So, in C_2 , in virtue of confused identity, B actually has a single specific referential intention that refers confusedly to both Jones and

²³Kripke’s own argument involves a distinction between ‘simple’ and ‘complex’ cases of singular reference (only in the 1977 version, however). But it is seemingly impossible to understand, from Kripke’s writing, what really distinguishes the two cases. Thus I have avoided the issue. But I hazard the following conjecture. The simple case is a case where A ’s utterance of d is the main piece of evidence A provides in the context for the intended referent of d on that occasion. The complex case, however, is just where there can be other equally important pieces of evidence, such as a non-linguistic demonstration, especially when the object in question is open to view.

Smith simultaneously. The conflict is within that intention, corrupting it, and not between intentions of two different types, one general and the other specific.²⁴

If interpretive contextualism is the way to go the same point can easily be made. For in that case the most reasonable or salient interpretation will also depend on two pieces of evidence—an utterance of a name and a demonstration of the referent—that are in fact conflicting. The ideal hearer should take everything in the context into account and this should result, in the case at hand, in two equally valid interpretations: either the referent is Jones, or the referent is Smith. At least as Kripke has described C_2 , the context does not make one interpretation objectively more reasonable than the other. Thus, it seems, contextualism does not predict that one of the interpretations should be the semantic one. If, on the other hand, the ideal hearer is assumed to be fully aware of the speaker's confusion, she should simply take the utterance to be corrupt and not yet available for evaluation as either true or false. And I don't see how the hearer can be 'ideal' if deprived of this crucial piece of information.

The general conclusion to draw from this case is that cases of confusion do not provide good theoretical reasons for positing new semantic categories or entities. Such entities are more plausibly introduced on the basis of cases where the speaker forms a referential intention without being confused about the object referred to (cases of 'edenic' reference, as it will be called in Chapter 3).

2.2.2 Reimer's argument against intentionalism

In a series of articles, Marga Reimer (1991a, 1991b, 1992a, 1992b) develops a theory of what is referred to by uttering demonstratives like 'this' and 'that' on a given occasion. She makes her puzzle-driven methodology quite explicit and states that 'abnormal' and 'atypical' contexts—these are invariably contexts with confused speakers—allow us to decide between competing theories when 'normal' contexts are insufficient (1991a: 197). Such contexts are, in her opinion,

²⁴Michael Devitt (1981b, 1981a: §5.4) presses exactly this point. On his view, there are causal designation-chains or networks ('d-chains') between the referent and what the speaker intends or has in mind. And it is only in virtue of such networks that one object rather than another is the speaker's object of thought on a given occasion: "Confusions like the present one lead to a network being grounded in more than one object. Because there are d-chains to both Jones and Smith, I would say that neither was the speaker's referent but each was his *partial* referent (using a notion borrowed from Field [1973])" (Devitt 1981b: 515, emphasis in original).

“natural test cases” for theories of demonstratives (1992b: 398). Reimer considers three relevant theories: *intentionalism*, *contextualism*, and *quasi-intentionalism*. Regrettably, she assumes and does not argue for the claim that they are equally plausible across normal contexts. Showing this is more fraught with theoretical issues of descriptive and explanatory adequacy than Reimer seems prepared to admit. But is she, perchance, right to say that our pre-theoretical intuitions about atypical cases can arbitrate between rivalling views? I don’t think so. But let’s prod the details.

Reimer’s arguments are, conveniently, tailor-made for our purposes. What she calls intentionalism is simply our intentionalism as it applies to demonstratives. It is the view, which Reimer traces to Kaplan (1978), that the referent of a demonstrative, when uttered in a sentence on a given occasion, is determined by the speaker’s ‘directing intention.’ A directing intention is, roughly, the intention to make the audience attend to some object by means of available linguistic and contextual cues. Her, contextualism, which she traces to Wettstein (1984), is also just the application of our own definitions of contextualism to demonstrative expressions. According to this theory, the referent of a demonstrative is determined “entirely by certain publicly accessible features of the context” (Reimer 1992b: 377). Reimer, however, argues that these “all or nothing” views are mistaken. The middle-ground is quasi-intentionalism according to which there is a role for both context and intentions in reference-determination.

Here I focus exclusively on Reimer’s case against intentionalism, as this seems to be her most compelling argument, and illustrates her puzzle-driven approach perfectly. Of course, I cannot assume that intentionalism, as a general thesis, is true when presenting my objection. The primacy of directing or referential intentions is exactly what is at stake. It turns out, however, that Reimer misstates the implications the intentionalist view and her objection can thus be resisted.

She uses three confusion cases to show that, as she puts it, an accompanying demonstration can override a conflicting intention to refer to an object with a demonstrative (1992b: 373; 1991a: 190–191). The key case is her most worked-out example.²⁵

Key case. Suppose I forget my keys in my (shared) office while on my way home. I return

²⁵The others are Kaplan’s well-known Carnap/Agnew case (see below) and an example involving two dogs, Fido and Spot (1991b: 182).

to the office, where I see my officemate sitting at the desk. I suddenly spot my keys ('keys A') on the desk, right next to my colleague's own keys ('keys B'). Intending to grab my own keys, I distractedly pick up the other keys while uttering,

(12) These are mine.

Now, according to Reimer, intentionalism predicts that in uttering (12) I *actually referred* to my own keys. I clearly *intended* to refer to my own keys, the keys I was looking at before I grabbed the other keys. But my own keys are still on the desk and I'm holding my colleague's keys. And I am clearly, according to Reimer, referring to the latter and not the former with 'These' in (12). To show that this is intuitively right, Reimer points out that my officemate can appropriately say to me, 'No, you're wrong. *Those* are not *your* keys; they're mine.' Intentionalism thus makes the wrong prediction and, it seems, quasi-intentionalism is vindicated: the speaker's intention to refer to an object with a demonstrative can be overruled by other features of the context, with the result that the actual referent is some other object.

In fact, however, Reimer is wrong to think that intentionalism simply predicts that by uttering 'These' on this occasion I *actually* refer to my own keys. Intentionalism, Reimer and I agree, states that if a speaker intends to refer to *o* in uttering a demonstrative, *o* will be what the speaker refers to in uttering the demonstrative on that occasion. But the key case involves a conflicting intention to refer and it is not entirely clear why Reimer takes the intentionalist prediction to be so clear-cut in such a case.²⁶ As we saw in Chapter 1, conflicting referential intentions corrupt the linguistic evidence, and the utterance *misfires*: even if it seems absolutely clear to the hearer that the speaker *referred* to one object (keys B) it does not follow that the referential intention was not conflicting and the evidence not corrupt. Indeed, the evidence is *misleading* in principle. To see this, let us look at the key case in more detail.

As with Kripke's example it helps to compare two distinct utterances at slightly different times. Let us suppose, then, that I actually muttered (12) to myself at the moment I spotted my own keys, i.e. keys A. Call this (m)utterance U_1 . Perhaps my colleague heard me, perhaps she

²⁶Note that this use of 'conflict' is importantly different from Reimer's own. Reimer has in mind a conflict between an intention and the context, whereby the latter can override the former.

did not. But, again, I mistakenly grab keys B and repeat (12) loud and clear a few seconds later. Call this utterance U_2 .

Intentionalism predicts, correctly, that I referred to keys A in uttering U_1 . Here there is no intentional conflict or corruption of the linguistic evidence. But things have changed when it comes to U_2 . By then I have acquired the false belief that the keys I was looking at before are identical to the keys I am holding in my hand right now. More succinctly, I believe falsely

(13) *that keys A = keys B.*

Undeniably, then, I confusedly *intend* to refer to *both* keys A and keys B by ‘These’ as it occurs in U_2 . But how can Reimer then claim that intentionalism predicts that I refer only to the latter? She seems to have smuggled in some contextualist assumptions here. Since all the evidence strongly favors keys B—I’m holding them in my hand as I speak—Reimer is inclined to accept it at face value. But my false belief (13) clearly makes the evidence misleading, since my intention is *also* to refer to keys A, the keys I distinctly remember spotting. Moreover, in support of this, intentionalists could invent their own corroborating rejoinder, just like Reimer. My officemate might say to me, while grabbing and showing me my own keys, ‘You mean *these*?’ To which I could respond: ‘Of course, that’s what I *meant*!’ But this only shows how flimsy this sort of data is as evidence one way or the other.

So intentionalism, it seems, is off the hook. There is no need to insist that our semantic theory must predict a single object as the referent of ‘These’ in U_2 . There are two distinct objects intended by the speaker and, in this particular case, one of them is intuitively much more likely than the other to be assigned as referent by the hearer. But why should we say that this is the *actual referent*? I can find no reason other than raw intuition. I agree with Devitt (2012a) that intuitions of this sort are theory-laden and thus, it seems, it is a theory-laden question what the intuition really tracks. Plausibly, however, it tracks the fact that keys B are much more likely than keys A to be assigned as referent by a rational and competent hearer. If so, the intuition does not help us distinguish intentionalism from interpretive contextualism. These theories are in total agreement about, first, what the most natural interpretation is (keys B) and, second, that the intention of the speaker does not uniquely determine the most natural interpretation (keys B) as the referent of ‘These’ in U_2 . So much for the natural test case.

Importantly, however, if the intentionalist theory is expanded to allow for some type of weighing of different (possibly conflicting) referential intentions, we may squeeze out some more definite predictions. And Reimer's (1992b: 388–396) own quasi-intentionalism leads in this reasonable direction by distinguishing *primary* from *secondary* intentions. Here is how this would work for confusion cases in particular. If I am confused in virtue of believing (13) but I form an intention to refer demonstratively to, say, keys B, *in a context in which keys A have no salience whatsoever*, then it seems like my intention determines keys B as what I refer to on that occasion of utterance. Suppose I mistakenly took my officemate's keys and went on vacation. I have had keys B in my pocket for over a week without using them—still believing implicitly that they are identical to my own—and I say, slightly annoyed, 'These are damn useless around here,' while putting the keys away somewhere. My primary intention is to refer to the keys in my pocket and my false belief seems pretty much irrelevant. This does, however, make the example significantly different from Reimer's own: there is absolutely no intuitive pull in the claim that I *really intend* to refer to keys A, still lying on my desk far away.

Reimer (1991a: 191–193; 1992b) alleges that Kaplan's much discussed Carnap/Agnew example and her key case lead to the same conclusion, i.e. that intentions can be overridden by context. Kaplan imagined a context in which he points to a picture on the wall behind him and says something like (14).²⁷

(14) That is a picture of one of the greatest philosophers of the twentieth century

Kaplan thought he was pointing to a picture of Rudolf Carnap, but unbeknownst to him the picture had been replaced by a picture of Spiro Agnew. Now, did Kaplan refer to the picture of Carnap that he 'had in mind' or to the picture of Agnew that he (appears to have) 'actually demonstrated'?²⁸ Agreeing with Bach (1987: 183–186) and Neale (2005: 182) I don't think there is much theoretical point in answering such questions in those terms. We have the (possibly confused) referential intention, and two possible interpretations of the utterance. No more

²⁷For a sampling of views about this particular example, see Borg (2004: 152–153); Heck (2002: 17); King (2013: 296–305).

²⁸The parenthetical remark is added because pointing itself is, most theorists appear to think, a thoroughly intentional notion. McGinn (1981) is a nice example of the problems engendered by ignoring this fact (cf. Wettstein (1984: 83–84)).

is required to answer the basic semantic questions this example raises. Anyhow, this case is even less compelling than the key case, because Kaplan's *primary* intention seems to have been to say something *about the picture of Carnap*. The fact that he is confused in virtue of believing, to put it simply, *that the Carnap-picture he remembers is the Agnew-picture he (appears to be) pointing to*, doesn't matter much, practically speaking, in the context. Furthermore, if the audience knows that Kaplan believes that he is pointing to a picture of Carnap—not altogether unlikely—they will immediately get Kaplan's primary referential intention. In that case this would be similar to mere slips of the tongue or malapropisms where everyone knows what the speaker actually intended. Malapropisms are discussed in detail in the next Chapter, §3.3.

Going contextualist doesn't change matters. Without going into the details, I would take issue with the common assumption that interpretive (or objective) contextualism, simply and unproblematically, predicts that Agnew is the referent of 'That' in (14). An ideal interpreter would, I suppose, be in possession of some very mundane facts about the speaker and the context which would make the Agnew-interpretation appear absurd. What possible reason could there be to allow the ideal hearer to have access to only a specific subset of facts, so as to make Agnew appear to be the right choice? Admittedly, it is easy to imagine someone who would interpret the utterance in this way. But this mere possibility of misunderstanding shows nothing.

To sum up, Reimer's argument relies on a confusion case, making predictions less clear-cut than she assumes. This undermines her claim that semantic puzzles are natural test cases for metasemantic theories like intentionalism and contextualism. The moral of the story is more generally that this kind of puzzle-driven argument should be rejected or, at least, needs sturdier foundations. Furthermore, and in tune with the objection to Kripke above, new semantic categories—in this case the category of 'actual referent'—are not credibly postulated on the basis of nothing more than judgments about puzzles involving confused speakers.

2.2.3 Perry on ‘automatic’ indexicals

John Perry has produced many influential arguments that fit squarely into the tradition of puzzle-driven semantics.²⁹ His famous argument for the ‘essential indexical’ (Perry 1979), based on a scenario where he unknowingly leaves a trail of sugar in a supermarket, arguably depends on doubtful assumption about the referential intention of a confused speaker. This particular argument has spawned a huge literature and many of the objections that have been raised are compelling (e.g. Devitt 2013b), so I will not focus on it here. Instead, I aim to establish that Perry’s distinction between ‘automatic’ and ‘discretionary’ indexicals is groundless. This issue is discussed in detail in the second edition of his book *Reference and reflexivity* (2012).

Remember the example from page 45 above. Rip Van Winkle falls asleep on July 3, 1766, and when he wakes up exactly twenty years later he utters (9).

(9) I fell asleep yesterday

Perry agrees that since Van Winkle believes he only slept a single night he *intends* to refer to July 2, 1766 by uttering ‘yesterday.’ According to Perry, however, he will in fact have designated July 2, 1786. On his account, the semantics of ‘yesterday’ are such that when uttered it refers by default to the day before the occurrence of the utterance.³⁰ ‘Yesterday’—along with ‘I’ and, perhaps, ‘now’ and ‘here’—are automatic indexicals in the sense that the speaker’s referential intention does not determine the referent of the expression on an occasion. Perry claims that the only intention relevant in this type of case is the speaker’s intention to be “using the words with their ordinary meaning” (2012: 70).

The view seems intuitive enough. First, one might think that, surely, a competent speaker can felicitously use ‘yesterday’ without knowing anything about yesterday save that it was the day before today. Secondly, this idea seems to mark a significant distinction between automatic and ‘discretionary’ indexicals. In the latter group we find, for example, complex demonstratives like ‘that man.’ On Perry’s theory, if I utter ‘That man stole my wallet,’ pointing to one particular

²⁹Perry (2012) contains so many confusion-based arguments and examples that I lost count, cf. pp. 70, 73–75, 77–78, 158, etc.

³⁰Thus Perry seems closer to objective contextualism than interpretive contextualism. Here, I treat him as an interpretive contextualist for sake of argument.

man in a crowd, the referent depends on my intention. As he puts it, “I have some choice or discretion in the matter” (ibid.).³¹

In my view, the Van Winkle example does not show that ‘yesterday’ is an automatic indexical in Perry’s sense. The case is *not* one where the speaker *simply* intends to refer to the day before today. Before arguing for this claim, however, it is important to explain why the example can deceive one to think otherwise. Van Winkle’s utterance in (9) is no good as data because the speaker has a conflicting referential intention, but also because the imagined context is too far-fetched to elicit reliable intuitions from competent speakers in the first place. Why on earth should we use such an absurd example when there is clearly a plethora of actual utterances involving the same semantic issue?

Accordingly, imagine a less fanciful context in which Eileen falls asleep on day 1 and—for whatever reason—sleeps all through day 2 and wakes up early on day 3. Believing that today is really day 2, Eileen utters (9).³² Now, which of the following is *more* accurate as a specification of the proposition she in fact expressed by her utterance?

(15) *that she fell asleep on day 1*

(16) *that she fell asleep on day 2*

Looking at this from the intentionalist point of view, one must take full account of Eileen’s identity confusions. She has many false beliefs when she wakes up, for example *that today is day 2, that yesterday was day 1, and that day 3 is day 2* (note that these are formulated in the theorists’ *unconfused* language, not Eileen’s own). These confusions could be verified if we supposed that Eileen had a calendar where she had marked day 1 in red before going to sleep, like this:

Day-1	Day 2	Day 3
-------	-------	-------

³¹Perry then goes on to use a confusion case to argue that it is only the so-called *directing* intention that really determines the referent. Since his argument for this claim is quite similar to Reimer’s, which I have already discussed, it is ignored here (but see also Perry 2009).

³²Note that ‘yesterday’ is an adverb and not a noun, and thus would not really be classified as a referring expression on some theories. This should be fine if we focus on cases where PPs containing explicitly referential NPs like ‘day 1’ can easily be substituted for the adverbs.

In the theorists' unconfused language, of course, today is really day 3 and yesterday was day 2. Clearly the intentionalist will have to say that Eileen's referential intention is conflicting and the evidence provided by 'yesterday' in (9) thereby corrupt. But one intention certainly seems primary and the other secondary. As with Reimer's example (12) above, this is partly explained by the rest of the utterance. For Eileen clearly intends to refer to a day with a certain property, namely the property of being the day she last fell asleep. This indicates day 1 rather than day 2 (contrast: 'Yesterday was a day'). The bottom line: intentionalism predicts that Eileen confusedly intended both (15) and (16), but can easily accommodate auxiliary assumptions about the relative weight of the intentions, even if they are in conflict.

But Perry may persist in his claim that (16) must describe the *actual* proposition expressed by Eileen. With this down-to-earth example, however, the suggestion loses whatever intuitive plausibility it had. Consider the fact, for instance, that Eileen's interlocutor could be fully aware of her confused mental state. Call him Frandie. Frandie could then point to day 1 on Eileen's calendar and ask, 'When you say "yesterday," do you mean *this* day?' To which Eileen would answer affirmatively if she is honest. It would be perverse to insist that Eileen actually meant and said (16) by her utterance. If (9) is uttered by Eileen in the witness stand while undergoing cross-examination by Frandie in a court of law, it's clear that Eileen will ultimately (i.e. once her confusion is cleared up) be taken to have meant that she fell asleep on day 1.

It may seem better for Perry to lean on interpretive (or objective) contextualism. No one denies that it is a rule, convention, or some sort of generic truth about English that 'yesterday,' in one of its lexical entries, *ought to* refer to the day before the utterance. Another lexical entry has it referring to the past more generally ('Yesterday's solutions are not good enough'). Assuming the former convention will suggest that (16) is the most natural interpretation of (9). But conventions, 'shoulds,' and 'natural' interpretation only gets you so far. The contextualist really faces the same issue as the intentionalist: when the ideal interpreter is aware of the speaker's confusion, the conventions of the language plus the context won't determine a single correct interpretation. Moreover, even if ideal interpreters are defined in such a way that they cannot be awake to this particular aspect of the speaker's mental state, it is quite unclear in what sense they will be able assign the right interpretation. Frandie has not understood

Eileen's utterance, as the context is described, if he simply thinks that she referred to day 2 by uttering (9). The same should be said about the Rip Van Winkle case by way of extrapolation from the more familiar to the less familiar.

Even the intentionalist concurs that uttering a word like 'today' or 'I' *almost* automatically refers to the day of utterance or to the speaker on account of their linguistically encoded meanings. But if the speaker is confused, special care is called for in describing the utterance and the meaning intended.³³

2.3 Donnellan on descriptions: Limiting case?

In this section I explain how there can be cases where the linguistic evidence is misleading or corrupt because the speaker has certain false beliefs but, since the referential intention itself is not conflicting, they don't actually fit the mould of puzzle-driven semantics. Take Kripke's (1977) reconstruction of one of Keith Donnellan's (1966) arguments for distinguishing between attributive and referential uses of definite descriptions. We are at a party, making small-talk, when I say to you

(17) The man over there drinking champagne is happy tonight.

I intend to refer to a certain man, call him Alfred, whom we can both see standing at the bar with a champagne glass. Unbeknownst to us, Alfred will be driving home tonight and the glass actually contains some nonalcoholic beverage that looks a bit like champagne.

³³I should note that the list of puzzle-driven arguments of the Kripke-Reimer-Perry variety is much longer, so this could go on *ad nauseam*. Here is a nice example from Brian Loar:

[T]he radical two-use theory [i.e. that singular terms are ambiguous between referential and attributive use] implies that a *sufficient* condition of understanding a referential utterance of '*t* is *G*' is merely correctly identifying the referent of *t* and the property expressed by *G*. But that is not sufficient. Suppose that Smith and Jones are unaware that the man being interviewed on television is someone they see on the train every morning and about whom, in that latter role, they have just been talking. Smith says 'He is a stockbroker,' intending to refer to the man on television; Jones takes Smith to be referring to the man on the train. Now Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train; but he has failed to understand Smith's utterance. It would seem that, as Frege held, some 'manner of presentation' of the referent is, even on referential uses, essential to what is being communicated. (1976: 357)

Assume, with Kripke, that there is exactly one man ‘over there’ drinking champagne. His name is Jeff and he is in fact miserable. We have no idea that Jeff is there. The main point of Donnellan’s classic paper was that when I utter (17) in this context it certainly seems like I have referred to Alfred and said something true of Alfred, namely that he is happy. This suggests that definite descriptions can be used *referentially*, i.e. as evidence for the speaker’s referential intention, as well as *attributively*, i.e. merely as evidence for some identifying condition specified by the description. If the description in (17) is construed attributively, it seems, what I said was about *Jeff* and *false*. This is of course counterintuitive.

It may seem like my critique of the puzzle-driven approach ought to apply to this argument just as well. Actually, however, it does not. Just like a confused speaker may produce linguistic evidence that is corrupt, speakers with false beliefs may produce evidence that is misleading. My false belief that the man I’m looking at is drinking champagne explains why my utterance is, though possibly true, misleading. Further, many argued in response to Donnellan that since the definite description doesn’t apply to the intended referent, i.e. Alfred—because it really applies to Jeff—we cannot maintain that *what was said* applies to the intended referent. Only *what was meant* applies to Alfred.³⁴ But the analogy goes no further than this.

There is a crucial distinction between this case and confusion cases. In the champagne example the speaker does not suffer from confused identity and his referential intention is, thus, as precise as can be. If it were a confusion case, the explanation of why my utterance is misleading would have to be my false beliefs about identity. But we don’t explain why (17) is misleading by saying that I implicitly believe something like (18) or (19).

(18) *that the man over there drinking champagne = Alfred*

(19) *that Jeff = Alfred*

The description in (18) must be interpreted as referring to Jeff *or* specifying a condition that applies uniquely to Jeff in the context of utterance. Thus (18) and (19) are stated in the un-

³⁴Thus Lockwood (1975: 486n21) writes that Donnellan “[rides] roughshod over the common-sense distinction between what a speaker means and what he actually succeeds in saying.” See, also, Bach (1987); MacKay (1968); Neale (1990); Salmon (1991); Wettstein (1981); Wiggins (1975). For arguments more compelling than Donnellan’s own in favor of the semantic significance of the distinction, see Devitt (2004) and Reimer (1998).

confused theorists' language. In the context of (17), attributing (18) or (19) to me as implicit beliefs is simply wrong. *Ex hypothesi*, neither I nor my interlocutor has formed any notion at all of Jeff. It is key to my case against puzzle-driven arguments that they essentially invoke examples of conflicting referential intentions, thus I do not object to Donnellan's argument. In uttering (17) my intention is clearly to refer to Alfred and not to Jeff, regardless of the fact that the description, in some sense at least, applies only to the latter.

If Alfred and Jeff are look-alikes and I utter (17), pointing to Alfred, while having mistaken one for the other regularly at the same party, then we might have a problem with constitutively determining the intended referent, and Donnellan's argument for a referential use would be considerably weaker.³⁵ In such a context, something like (18) or (19) could have been appropriate for explaining why (17) was misleading. In the next section I discuss one last example of a flawed puzzle-driven argument, but this time the topic is anaphora and coreference rather than singular reference as such.

2.4 Interlude on puzzles and anaphora

Scott Soames (1989c, 1989b, 1994) develops sophisticated puzzle-driven arguments against, as I shall call them, *inheritance-theories* of the semantics of anaphora with c-commanding singular term antecedents. According to inheritance-theory, anaphors inherit their interpretation from the interpretation of their antecedents. There are two varieties, Millian and Fregean. Soames argues against both but, for simplicity, I focus only on the Millian here. On this view, anaphors inherit the *referents* of their antecedents and, further, the referents exhaust the semantic contents of the anaphors in a given utterance.

To be clear, the purpose of this section is, first, to show how one very influential argument against inheritance-theories of anaphora fails, not to argue that such theories are actually correct. And, secondly, to show how the argument fits the mould of puzzle-driven semantics. In fact, Soames' alternative proposal is much more plausible as a theory of the semantics of anaphora.³⁶ As it turns out, however, applying our model of confusion to cases involving

³⁵On the distinction between constitutive and epistemological determination, see Neale (2005: 180).

³⁶Soames' (1994: 119) theory is that occurrences of pronouns anaphoric on c-commanding

coreferential constructions gives rise to some interesting constraints on the proper function of such constructions. Those constraints are developed and defended in Chapters 4 and 5.

Now, according to Millianism (for short) the sentence,

(20) John loves his mother,

contains two occurrences of directly referring expressions, 'John' and 'his.' Since the occurrence of 'his' is assumed to be anaphoric on the antecedent occurrence of 'John' it simply refers to the same thing as 'John' in (20) and the referent is its semantic content. Millianism, according to Soames, identifies the proposition ('semantically') expressed by (20) with that expressed by 'John loves x 's mother' relative to an assignment of John to x (1994: 115).

Soames then introduces a confusion puzzle and argues that the Millian theory cannot explain it. Thus, as he puts it, the theory is unable to account for "the full range of cases that need to be covered" (ibid.). Here is how Soames sets up his example. He supposes that the sentence,

(21) Mary says, and believes, that John loves his mother,

is true in virtue of Mary having uttered (20) while pointing at John. Soames points out, however, that on the Millian picture the truth of (20) and (21) doesn't require that Mary *realizes* that she is pointing and referring to John himself while uttering 'his' in (20). We can assume, even, that Mary is convinced that the person she refers to is *not* John. Either way, she fits the model of confusion argued for in Chapter 1. Thus, Soames continues, the truth of (20) and (21) doesn't require that Mary say or believe that John is an own-mother-lover. But this comes into conflict with at least one anaphoric interpretation of (21) because (21) *can* be uttered to express the proposition that Mary believes a proposition attributing to John the reflexive property of loving one's own mother. But then, according to Soames, there is at least one legitimate interpretation of (21) that is not captured by Millianism, namely the reflexive interpretation (ibid., 116). Thus,

singular terms must be construed on the model of bound variables. Either singular terms are really (restricted) quantifiers that bind their anaphoric pronouns or the anaphoric relation introduces a λ -operator which then binds the anaphoric pronouns. On the first account (20) has a (reflexive) reading on which it is analysed as [John x](x loves x 's mother). On the second, (20) has a (reflexive) reading on which it is analysed as λx (x loves x 's mother)John.

Millianism is inadequate.³⁷

If this interpretation of Soames' argument is correct—and I struggle to find a more plausible construal—it's clearly no good.³⁸ First, there is no *conflict* here at all: /his/ is the superficial form of at least two distinct pronominal expressions. One expression is reflexive and the other is non-reflexive.³⁹ Soames is imagining a case where 'his' in (20) is non-reflexive (and *deictic*) and 'his' in (21) is reflexive. These are simply two different expressions and thus it can be argued that (21) isn't really an accurate report of what Mary said when she uttered (20). It might still be true, strictly speaking, to report Mary as *believing* the proposition that John is an own-mother-lover—she did point to John. She just doesn't fully realize this and certainly did not *intend* to express this proposition. Secondly, the Millian theory of anaphora doesn't make any prediction about the behavior of deictic non-reflexive occurrences of 'his.' This shouldn't be surprising since the content of a pronoun which is accompanied by a pointing gesture must be determined by something like what the speaker intends to point out in the environment. In these occurrences it behaves more like demonstratives, such as 'this' and 'that' and less like antecedent reference inheritors.

Lastly, and most importantly, the context described by Soames is incoherent. Mary cannot intend to use a reflexive form of 'his' and, at the same time, intend to indicate the referent of 'his' by pointing—but this is required if (21), with a reflexive 'his,' is to be a fully accurate report of the utterance in (20). Either the pronoun is reflexive and the competent speaker therefore intends 'his' and 'John' to indicate the same object or the pronoun is deictic and non-reflexive

³⁷Fiengo & May (1994: 5–6) discuss a very similar case but reach a conclusion consonant with my own below.

³⁸To be fair, Soames argues that this failure of the Millian theory can be fixed by switching to a Fregean alternative, according to which the speaker cannot assert the reflexive proposition without recognizing “the purported lover and the person whose mother is said to be loved as one and the same” (ibid., 115). But the Fregean theory then faces a very similar puzzle to which my arguments would also apply. For simplification I focus on Millianism here.

³⁹See, e.g., Fiengo & May (ibid., ch. 1); Neale (2005: 275). According to binding theory, as propounded by Fiengo & May, in one case the pronoun and subject NP are coindexed and therefore coreferential in virtue of linguistic meaning. In the other they are noncoindexed and, so, can be either coreferential or noncoreferential—both options will be compatible with the linguistic meaning of the sentence relative to the index. According to Neale, the non-reflexive pronoun can be further divided into two distinct types: those that are bound and those that are unbound. This distinction doesn't matter for our discussion, so I ignore it here.

and the competent speaker is not required, by the semantics of English, to be intending to indicate the same object twice. But if the pronoun is really deictic it cannot be anaphoric. If so, Soames has not described a context that a theory of anaphora needs to capture.⁴⁰

These points can be clarified more readily by changing the object language into one that lexicalizes this particular pronominal ambiguity. Icelandic is one such language. In Icelandic (20) has two literal translations.

(22) John elskar móður hans [*non-reflexive*]

(23) John elskar móður sína [*non-deictic, reflexive*]

But (23) would be unacceptable in Soames' context unless Mary is, in tandem with the occurrence of 'sína,' pointing to an x such that she takes x to be *the mother* of the person she intends to refer to with the occurrence of 'John' in (23). The only other option is that she does not point to anyone at all—the expression 'sína' is non-deictic and obligatorily reflexive on its own.

Further, and this is corroborated by Thráinsson's (2007: 461–462) work on Icelandic syntax, (22) is also unacceptable if Mary intends 'hans' and 'John' in (22) to be coreferential.⁴¹

Thus, a competent speaker of English cannot intend to use (20) reflexively and non-deictically in this context while also intending the reference of 'his' in (20) to be indicated by her point-

⁴⁰In a single sentence, Kit Fine (2010a: 496) anticipates exactly this line of argument in his response to a Lawlor's (2010) criticism of semantic relationism: "I suppose it might be doubted that a semantically competent speaker could use the pronoun in this way." Characteristically, however, Fine is quick to reject this by improvising a decidedly fanciful counterexample:

I am hosting a party and two of my guests shout, 'John is here,' having a different person in mind. I hear the two utterances of 'John' as one utterance and respond by saying 'I hope he brought some wine,' taking my use of 'he' to be anaphoric on what I take to be one utterance of the name. There is then confused reference, I assume, but no semantic incompetence. (ibid.)

Without some specific reasons, I am at a loss why this odd case needs to be included within the purview of our theory—hearing two utterances *as one*?—and since Fine provides no such reasons further comment is gratuitous.

⁴¹Thráinsson goes so far as to say that this would be ungrammatical: "Hence it is fine to say *Egil took his book*, with coreference (coindexing) between *Egil* and *his*, but the corresponding sentence is ungrammatical in Icelandic [...] as the reflexive possessive has to be used" (ibid., 262). But such a strong claim isn't needed for the present argument.

ing to the one whose mother she has in mind. Therefore, Mary seems to be using (20) non-reflexively in the context in question, but we were only interested in its reflexive interpretation.

But what if Mary is simply *indifferent* as to whether she refers to the same individual twice or to two distinct individuals? Suppose that Mary, now speaking Icelandic, utters (22) and unwittingly and indifferently refers to John himself with the occurrence of ‘hans’ (maybe she’s blindfolded). Now the report corresponding to (21) in Icelandic can either contain ‘hans’ or the obligatorily reflexive ‘sína.’ But both reports seem to be false or, at least, misleading. Mary’s indifference has the effect that she doesn’t know what proposition she (was taken to have) expressed. So she neither intended to say, of some x , that x loves x ’s own mother, nor did she intend to say, of some x and some y , that x loves y ’s mother. And she doesn’t believe either of these propositions. The issue of referential indifference is discussed in detail in Chapter 4. Suffice it to say, here, that Soames’ argument is not helped by assuming indifference.

Now consider two variants, V_1 and V_2 , of Soames’ own context, such that Mary is confused in neither variant and intends ‘John’ and ‘his’ in (20) to corefer in both. In V_1 Mary uses ‘his’ in (20) non-deictically since John is nowhere to be seen. And in V_2 she uses ‘his’ deictically, believing truly that she is pointing to John himself, if such a context is plausible at all (one might think such a situation forces a reflexive use).

Arguably, the only linguistic difference between V_1 and V_2 is that the ‘his’-tokens belong to different linguistic types, although the difference is not audible like in the Icelandic translations (22) and (23). In terms of these non-confused utterances, Soames’ case gives the Millian no reason to deny that the proposition Mary intends to express can be exactly the same in both variants. Namely, the proposition that Mary is an own-mother-lover. She is only using slightly different linguistic devices to communicate this proposition. The two devices are identical in (English) surface form, but differences in the perceptual environment explain why Mary will have slightly different communicative intentions in V_1 and V_2 .

Perhaps the following example provides a more intuitive case for Soames’ argument. But it still fails. Think of a context, C , in which Lois Lane believes falsely that Superman is not identical to Clark Kent. She utters (24) and points to Clark Kent in tandem with ‘his.’

(24) Superman loves his mother.

Now, what should a semantic theory assign as content to (24) in *C*? I don't know—but my more general point is that this is simply a non-issue. Lois clearly *intends* to express a proposition about two individuals but fails to do so. Does she therefore express a proposition about a single individual? Well, that's clearly not what she intended. And positing such an unintended interpretation, in the case at hand, would require serious argument and independent motivation. On the face of it, however, such a position appears *ad hoc*.

Turn briefly to contextualism. What exactly would the ideal contextual interpretation of (24) be? I'm not sure that this question has a good answer either. Either the ideal interpreter (i) is aware of Lois' confusion or (ii) he is also confused and believes that Superman is not Clark Kent. But in (i) the interpreter would know that the proposition intended isn't even possible, because it was supposed to be about two distinct individuals, and, if speaking Icelandic he could offer a grammatical correction. And (ii) is hardly an example of an ideal interpreter. Further, if Lois realizes that she (could be understood as) referring to Superman *twice*, she won't think that she actually intended to express the proposition that Superman is an own-mother-lover. She intended no such thing, so it makes no sense to say that this is what she 'really' expressed.

Indeed, if the reflexive and non-reflexive readings involve two different types of pronoun, and we assume that the speaker's intentions determine both which pronoun is uttered and the content of the utterance, then it becomes impossible to construct a plausible context in which the speaker utters (20) with a non-reflexive 'his,' intending thereby to express a coreferential proposition.

The moral is that identity confusion can make an otherwise intelligible context involving (apparent) coreference into something internally incoherent. More positively, this suggests, or so I argue, some cognitive constraints on genuine coreference, which will be discussed in Chapters 4 and 5.

Conclusion

Summing up, there is good reason to think that the definition of puzzle-driven semantics actually characterizes a substantial amount of what passes as philosophy of language today. Unfortunately, the basic theoretical goal of this research program is neither worthwhile nor

does it offer any clear contribution to the task of explaining linguistic behavior more generally.

Philosophers in this tradition have, moreover, tended to employ confusion-type puzzles in support of theories about the semantics of specific linguistic expressions, especially expressions designed to enable singular reference. These are arguments that either (i) appeal to assumptions (or ‘intuitions’) about what speakers refer to in making confused utterances, and/or (ii) aim to establish negative or positive conclusions about the exact contents expressed by confused speakers. If the general direction of my own reasoning is right, the arguments are mistaken, and mistaken because of fundamental flaws at the very core of the research program.

It is high time to present, in more detail, the alternative view of semantics as explanation-driven, and to show how the basic semantic questions informing this different perspective can be answered. This is the burden of Chapter 3. In the next Chapter, however, I focus on a prominent puzzle-driven argument against a certain view of *anaphora* and *coreference*.

Chapter 3

Arguing for Gricean intentionalism

Introduction

I have argued that philosophy of language needs to be explanation-driven in the sense that its overarching theoretical task or aim is to explain the fact of successful communication. But how should we go about fulfilling this task? And, in particular, how do we explain successful acts of referring with a singular term? The rest of the dissertation aims to address these two questions directly, based on conclusions already reached. My answer will be grounded in the tradition of Gricean intention-based semantics and pragmatics.

In this Chapter, I present the basic intentionalist framework, defining core notions like speaker meaning and communicative intentions (§3.1). Without going into much detail, I present a bare-bones theory to provide the basis and background for what was called intentionalism in the last Chapter (p. 43): the basic commitment that what is said by a speaker in making an utterance is determined only by the speaker's communicative intention. This also embodies the commitment that what is referred to is determined only by the speaker's referential intention.

Subsequently, I present a two-pronged defence of the intentionalist program. First, in §3.2, I argue that the phenomenon of underspecification—the thesis that linguistically encoded meaning always underdetermines the meaning intended by a speaker on an occasion—provides support for the program. Secondly, in §3.3, I argue that a common type of counterexample to the Gricean notion of what is said fails. This is the case of *malapropisms*. Just like confusion-based puzzles, malapropisms have been used as counterexamples to intentionalism about what is said. And just like the former, they fail as such.

I present two arguments against this traditional construal of malaprops; the first is based on prior results about underspecification, the second is independently motivated and charges that the objection is committed making a scientifically arbitrary distinction. I also present my own positive theory of malapropisms, which is consonant with the general Gricean program presented here, and draws support from experimental work on speech errors by phoneticians and phonologists (§3.3.2).

3.1 Gricean speaker meaning

According to Gricean intention-based semantics speaker meaning can be defined in terms of a special kind of audience-directed intention made up of, at least, two sub-intentions. To illustrate, consider the following example, adapted from Michael Tomasello (2008: 90).¹

Yet another drink

Somehow I'm stuck at this pompous, formal-attire party and my choice of outfit is slightly off the mark, or so I come to think (I'm the only one wearing black *jeans*). My boss from work is hosting and serves the drinks. I'm deeply embarrassed by how quickly I drained my first glass, but I'd still like a refill as soon as possible. Thus,

- (i) I decide to place my empty glass at a perspicuous location without anyone noticing, hoping, correctly, that my boss immediately fills it up without recognizing that it's *my* glass.

Summoning all the self-control I can muster, I spend more time on my second glass. When I desire another drink,

- (ii) I make eye contact with my boss and, nodding suggestively, look down at my empty glass which he readily fills to the brim.

When I want *yet another drink* my inebriated state has erased all norms of etiquette from memory and

¹His example is based on Sperber & Wilson (1986/1995: 46–50). See also Wilson & Sperber (1993: 150–152).

- (iii) I say loudly, ‘Another drink,’ whereupon the host finds my glass—the only empty one on the table—and fills it with water.

On the Gricean way of seeing things, this example involves two *utterances*, one in each of (ii) and (iii).² The latter is a *linguistic* utterance, the former a non-linguistic and, arguably, unconventionalized act of communication. In (ii), I intend for the host to recognize that I want another drink and I even intend, further, for the host to recognize *that I intend him to recognize my desire for drink*. The first case (i) is very different in that it neither counts as an utterance nor as an act of intentional *communication*. Yet in (i) I have an *informative* intention of a certain sort: I intend to have a certain effect on the mental states and, ultimately, behavior of my boss. But I wish to hide this intention of mine from him.

Simplifying considerably, and appropriating ideas from relevance theory, let us define the minimal notion of Gricean speaker meaning thus:³

Gricean speaker meaning (GSM)

Speaker *S* means something by uttering *U* iff there is an audience *H* such that *S* utters *U* intending

- (1) to produce thereby in *H* a certain response *r*;
- (2) *H* to recognize *S*'s intention (1).

Call (1) *S*'s *informative* intention and (2) *S*'s *communicative* intention.⁴ The latter is a reflexive intention to make *H* recognize that one has a certain informative intention. They are jointly necessary for acts to count as communicative utterances. In our story above, my action in (i)

²I follow Grice (1989: 118) in using ‘utter’ and ‘utterance’ in an “artificially wide sense, to cover any case of doing *x* or producing *x* by the performance of which [the speaker] meant that so-and-so.”

³See Grice (1957). Grice’s definition of speaker meaning has long been the subject of intense debate, cf. Bach & Harnish (1979); Grice (1969a); Neale (1992); Recanati (1986); Schiffer (1972); Searle (1969); Sperber & Wilson (1986/1995); Strawson (1964).

⁴Following Sperber & Wilson (1986/1995: ch. 1, unpub. ms.), among others, the definition skips Grice’s third condition, stating roughly that *H*’s recognition that *S* intends (1) should function, in part, as a *reason* for (1). The condition may well be on to something important about the nature of speaker meaning, but it brings with it particular complications that I wish to sidestep here. See in particular García-Carpintero (2001: 102–103); Neale (1992: 547–549); Wharton (2009: ch. 2).

only satisfies (1) in GSM but both utterances in (ii) and (iii) are *bona fide* examples of GSM. On this model, successful communication requires the speaker to have a great deal of true beliefs about the epistemic situation of the audience or hearer. For example, in (iii), *S* must believe truly that the hearer understands a specific English sentence and that, uttering ‘Another drink’ in this context, will cause the the hearer to recognize *S*’s desire for more drink.

In each of (i)-(iii), *S* intends to communicate a *request*. The act of requesting can be analyzed as a species of GSM:

Gricean request (GR)

Speaker *S* requests *H* to *V* by uttering *U* iff there is an audience *H* such that *S* utters *U* intending

- (1) to produce thereby in *H* the intention to *V*;
- (2) *H* to recognize *S*’s intention (1).

A similar definition can be given, the Gricean assumes, for all kinds of different utterance types with different illocutionary forces, e.g., ordering, asserting, supposing, asking, and so on. In GR, *V* stands for some action and, applied to our story above, *V* is the same for (i)-(iii), namely *to fill S’s glass*.

The basic hypothesis of Gricean intentionalism is that speaker meaning, as here defined, is explanatorily basic in communication. This is based on Grice’s insight that recognition of communicative intention is the basic task of the hearer, giving in effect a *sufficient* condition on successful communication.⁵ For one thing, the model provides resources to explain the intuitive continuity between linguistic and non-linguistic, or conventional and non-conventional, instances of communication. The former is prior in both the metaphysical and the historical-evolutionary sense. However, conventionalization and grammaticalization vastly enriches the speaker/hearer’s ability to express and entertain new beliefs, desires, and thoughts. Still, the basic mechanism of communication is grounded in the shared human capacity for forming

⁵See Sperber & Wilson (1986/1995: 25) on Grice’s “greatest originality” as they put it. I will take issue with this formulation, arguing that the cognitive constraints placed on the speaker are just as important for an intentionalist model of communication. In particular, see p. 108 in Chapter 4 below.

complex communicative intentions and inferring such intentions on the basis of observable behavior. More succinctly, it is grounded in the human capacity for metarepresentation (Scott-Phillips 2015; Sperber 2000: 121–127; Wilson 2000).

It is not to my purpose to develop a particular theory of how this account can explain the meaning of linguistic expression types in natural language. But many theorists have suggested different ways to go here.⁶ One possibility, for example, would be to follow Daniel Harris (2014) and explain the public meaning of conventionalized linguistic items such as ‘Another drink’ in (iii) in terms of overlapping communicative dispositions in a community. Speakers of English are disposed to utter ‘Another drink’ intending thereby—primarily, at least—to produce in the hearer an intention to give them another drink. Hearers have the overlapping disposition to recognize or infer that the speaker, in uttering this sentence type in the context as described, has this very intention.

Griceans assume, also, that GSM divides exhaustively into *what is said* by the speaker and *what is otherwise conveyed* or *implicated* on a given occasion of utterance.⁷ Both notions are defined in terms of the speaker’s intention, which is severely constrained by the speaker’s doxastic state, for example the speaker’s beliefs about the hearer’s own beliefs. Generally, one cannot form the intention to *V* at time *t* unless one lacks the belief, at *t*, that it is impossible for one to *V* at *t*.⁸ However, as Grice himself puts it, what a speaker *says* by uttering a sentence is “closely related to the conventional meaning of the words (the sentence) which he has uttered” (1975: 25). What is said, or the proposition(s) expressed, is thus constrained by a compositional semantics, i.e. the lexical meanings of the immediate constituents of the sentence and their syntactic mode of combination (see Unnsteinsson 2014). Implicatures are clearly not so constrained. Nowhere does Grice claim that what is said is fixed or determined by

⁶See, in particular, Grice (1989: ch. 6); Loar (1986); Neale (1992); Schiffer (1972, 2003). Many theorists believe—often based on Schiffer’s book *Remnants of meaning* (1987: ch. 9)—that the program is doomed and beyond repair (cf. Lance & O’Leary Hawthorne 1997: 287–297). But this is quite far from obvious.

⁷Grice (1989: ch. 2, 5, 6); Neale (1990: ch. 3, 1992: 523–524). But see Jennifer Saul (2002), where she argues, unsuccessfully in my view, that on the Gricean view “speakers may mean things which they neither say nor implicate” (p. 229). Trivially, she may be right about non-linguistic or non-conventionalized utterances (see note 12 below on this).

⁸See footnote 25 in Chapter 1.

conventional meaning: there only needs to be some type of close correspondence between what is said and the linguistically encoded meaning of the sentence uttered.⁹

Let's call the category of acts of saying 'assertives'—encompassing what a speaker states, claims, or says—and define it in this rough and ready way:

Gricean assertives

Speaker *S* asserts *that p* by uttering *U* iff *U* is an utterance of expression σ such that the linguistic meaning of σ is compatible with *p* and there is an audience *H* such that *S* utters *U* intending

- (1) to produce thereby in *H* the belief *that p*;¹⁰
- (2) *H* to recognize *S*'s intention (1).

For assertives the intentionalist assumes that *what is said* (or the proposition(s)¹¹ expressed) divides, at least, into (a) *what is referred to* and (b) *what is then predicated of what is referred to*.

⁹Thus, the relevance-theoretic notion of 'explicature' is supposed to do the work done in Grice's theory by 'what is said.' So, I disagree when Wilson & Sperber (2002b: 77) suppose that the latter is construed, by Grice and his followers, as an "intuitively clear, common-sense notion." Grice repeatedly refers to his own "favored" and "maybe in some degree artificial" sense of 'saying' (1989: 25, 33, 41, 86, 87, 88, 118, 121) which is clearly supposed to be a theoretical notion rather than a creature of mere common-sense (cf. Neale 1990: 64–65). The notion can be specified, roughly, as the coincidence of speaker meaning and linguistic meaning: the speaker says or asserts *that p* by uttering *U* when *p* is meant by the speaker and is uniquely compatible with the abstract linguistic meaning of the sentence uttered in *U*. See the definition of 'assertives' to follow.

¹⁰Grice (1969a: 106–112), of course, gave counterexamples to (1). In an oral history exam, the examiner asks the student when the Battle of Waterloo was fought. Student replies '1815' *asserting* thereby, it seems, *that it was fought in 1815*. But of course the student does not intend to produce this belief in the examiner, for he (the student) knows that she already knows *that it was fought then*. Thus, Grice distinguishes between *exhibitive* and *protreptic* utterances. For the latter (1) is left unchanged, for the former it becomes: to produce thereby in *H* a belief *that S believes that p*. McDowell (1980: §5) and Neale (1992: 545–547) point out that this makes our (exhibitive) utterances about our own mental states rather than about the world itself, which appears unintuitive. Based on recent work in relevance theory on hearers' epistemic vigilance—especially Sperber et al. (2010)—I incline to the view that this is entirely unproblematic. All our utterances could be exhibitive but we could still provide a rich explanation of how communication provides speaker/hearers with knowledge of the world. I keep the protreptic form here only for simplification.

¹¹I assume here, with many intention-based semanticists, that a single utterance can be intended to express multiple propositions, as well as multiple implicatures, cf. Bach (1999); Carston (2002: 125–134); Ciecierski (2009); Korta & Perry (2011: ch. 8); Murday (2014); Neale (1999, 2001); Wilson & Sperber (2012).

For convenience, since I am not particularly concerned with non-sentential or non-linguistic utterances, I assume here that all assertives involve an utterance in a natural language.¹² Both (a) and (b) are determined intentionally, (a) by *S*'s referential intention and (b) by *S*'s predicative intention (see next Chapter).

As indicated above, the speaker's saying-intention is constrained by her beliefs (some of which concern the expression used). If *S* believes that it is possible, say, that *H* will understand *S*, when *S* utters

(25) Troy was devastated,

on a given occasion, as referring to their acquaintance Troy—who is in emotional turmoil—and not to the Troy of Homer's *Iliad*, then *S* can *ceteris paribus* form that specific referential intention in uttering (25). Thus it is safe to assume that, by uttering (25), *S* said and meant the proposition *that Troy [the friend] was devastated [emotionally]*. And his utterance counts as successful communication if *H* understood that this was (part of) what *S* intended. This account meshes perfectly with the idea, explained in Chapter 1, that the communicative proper function of a singular term is to provide the hearer with evidence for a referential intention.

These are, of course, the mere beginnings of a full-blown intentionalist theory. But why should we believe any of it? I'll mention four reasons here. First, Gricean speaker meaning stands a good chance of being basic in the order of semantic explanation.¹³ Second, the lin-

¹²On this definition of assertives it would be impossible to *assert* or *say* anything without doing so by uttering some linguistic expression (or something *with syntax*, cf. Grice 1989: 87). And, to illustrate, in *Yet another drink*, it seems wrong to ask what was asserted in (ii), but *not* wrong to ask what was *meant*. Some seem to think this gives rise to problems, as the intentionalist surely needs non-linguistic acts of communication to ground properly linguistic acts (e.g. Harris 2014: ch. 1). But, to this purpose, it is enough to claim that the content of what is meant must be constitutively determined by the speaker's communicative intention, whatever its external manifestation. Thus, this is compatible with the idea that an act of communication doesn't count as assertion until it reaches the level of conventionalized or standardized forms of linguistic expression: one can still make a distinction between what is meant *directly* and what is meant *indirectly*, mirroring the saying/implicating distinction, for non-conventionalized means of communication. So one might have something like a distinction between proto-saying and proto-implicating, comprising the whole significance of a proto-assertive such as, in (ii), the act of showing someone one's glass while nodding suggestively.

¹³That is to say, any other notion of 'meaning,' ultimately needs to be cashed out in terms of speaker meaning to earn its theoretical keep. The idea is developed further in my "Compositionality and sandbag semantics" (2014: §6). See also note 12 in Chapter 2.

guistically encoded meaning of a sentence σ in some language L radically underspecifies the content intended by a speaker in uttering σ on a given occasion. And this fact should be seen as evidence for the intentionalist view. Third, the popular idea that speakers can, by uttering U on some occasion, say *that p* without also meaning *that p*, should be rejected. As we saw in §2.2, there is good reason to suppose that there is no reference without referential intention and thus, arguably, no saying without saying-intentions of the Gricean kind, either. Fourth, the minimal Gricean account of what constitutes communicative success, i.e. the idea that understanding an utterance is a form of mindreading or intention-recognition, is intuitively correct and has been quite fruitful both theoretically and experimentally. This last claim is examined further in Chapter 4.

In what remains of this Chapter, I focus on the second and third items on the list. First, I develop an extended underspecification-based argument for a broadly Gricean view of language and meaning (§3.2). Secondly, I drive the final nail in the coffin of unintentional sayings. Just like confused identity has been used to argue that intended reference and actual reference can come apart, *malapropisms* and slips of the tongue have been thought to provide for a very similar manoeuvre (§3.3). One of my argument against the latter is based directly on the conclusions about underspecification, so the issues are closely related.

3.2 Argument from underspecification

A major question in the philosophy of language is the following: what exactly is the relation between the meaning of a sentence and what is said by a speaker who utters that sentence on a given occasion? Traditionally the meaning of a sentence is identified with the proposition it semantically encodes relative to a contextual assignment of values to indexicals and similar devices, and *this* is thought to be the proposition expressed by the speaker, i.e. what the speaker strictly says. But this idea has faced serious challenges since the 1980's.¹⁴ The challenge can be stated in the form of a thesis, as follows.

(T) The linguistic meaning of an expression σ in a natural language L always constitutively

¹⁴See in particular Sperber & Wilson (1986/1995).

underspecifies the proposition expressed by a speaker who utters σ .¹⁵

Independently of my immediate concerns in this chapter, i.e. to argue for intentionalism, it is of considerable consequence whether (T) turns out to be true or false. If theorists are interested in explaining how communication is possible and what the role of language is in communication, (T) implies that explaining differences in propositional content in terms of differences in linguistic meaning will always fall short of full explanatory adequacy. And this is not just because of sentences containing indexicals and demonstratives, for (T) is a claim about all sentences in L . Here, I will explain five important examples of underspecification; each contributes to establishing the plausibility of (T). But I only take the last, i.e. underspecification from illocutionary force, as conclusive.¹⁶

3.2.1 Indexicality

The linguistic meaning of a sentence containing an indexical or demonstrative does not uniquely determine a referent for the indexical on a given occasion. The meaning of ‘he’ does not fix which male individual is being referred when uttered. The only meaning the pronoun has is that its referent must be male. And this, by itself, is not sufficient for a speaker to successfully refer to only one of two salient males. Thus a speaker normally assumes that the intended male is salient enough for such a minimal indication to be sufficient. Looking at the massive literature on this topic it seems clear that there is a growing consensus towards intentionalism. Reimer’s quasi-intentionalism—in §2.2.2, for example, was clearly a step in this direction.¹⁷

¹⁵Note that (T) is an empirical hypothesis about learnable natural languages, not a thesis about every conceivable language, and does not entail that a language which violates (T) is conceptually incoherent. On conceptual versus empirical claims in the philosophy of language, see Neale (2008: 390–393).

¹⁶Thus, to be clear, the notion of what is said at play here is a notion that *incorporates* illocutionary force. This is as it should be, as the competent hearer must recognize this part of the speaker’s intention if communication is to succeed. Needless to say, theorists have proposed many different ways of cutting the saying-pie. In particular, Bach (e.g. 1994a) defines a notion of a proposition type relativized to narrow context, which explicitly excludes creatures of illocution. I can safely ignore this here, as (T) is not intended as a claim about the underspecification of Bach’s what is said by encoded meaning.

¹⁷To take a few more examples, see Evans (1982: ch. 9); Kaplan (1989a); Kripke (1977); Neale (2005); Siegel (2002).

3.2.2 Ambiguity

Sentences containing ambiguous words such as ‘bank’ and ‘poker’ will underspecify the content of what is said on a particular occasion. ‘Bank’ is uttered either to mean a financial institution or a river bank. The meaning of say ‘Zoe is at the bank,’ by itself doesn’t constitutively determine which of the two contents is right when it is uttered. This doesn’t mean that both interpretations are true; when the speaker actually meant the river bank any other interpretation will be mistaken.¹⁸ Then there is syntactic ambiguity: ‘Flying planes can be dangerous.’

On some theories of proper names, they pose an identical problem: the occurrence of ‘Zoe’ in the above sentence goes no way, by itself, towards determining one and only one individual from the pool of people, fictional or non-fictional, bearing that name. This is because its encoded meaning is only that it should refer to *some* person bearing that name.¹⁹ This is a question of what *constitutively* determines the referent of ‘Zoe’ when uttered by some speaker. There is an analogous epistemological problem for the hearer, namely that knowing the meaning of the name type ‘Zoe’ is *insufficient*, by itself, for him to identify the Zoe in question.²⁰ This is, of course, a controversial issue since some theorists think names are individuated by their bearers: if such a theory is right, we simply have a case of ambiguity similar to that of ‘bank’ or ‘poker.’²¹ Yet the epistemic position of the hearer ought to be the same either way.

If all sentences in *L* are indexical or ambiguous, it seems like meaning could underspecify content in *L* with respect to all utterances on all occasions. But what if there are non-indexical and non-ambiguous sentences in *L*? Luckily for the proponent of (T) there are other kinds of underspecification.

¹⁸This does depend on how words are individuated, and I am assuming ‘bank’ to be the same expression type throughout. If this is thought unreasonable, the same argument can be given in terms of the phonetic sequence /bank/.

¹⁹For theories of broadly this type, see, e.g. Bach (1987: ch. 7); Elbourne (2005: ch. 6); Fara (2015); Recanati (1993: ch. 8). Similarly, Perry (2012: §6.3) argues that giving something a name establishes a *permissive convention* such that speakers are from then on *allowed* to use that name to designate the person in question. This results in a special type of ambiguity for names which, in Perry’s estimation, warrants the neologism ‘nambiguity.’

²⁰On the distinction between constitutive and epistemological determination, see Neale (2005: 179–181, 192–193). Also Devitt (2010, 2013c: §2).

²¹For arguments, see Devitt (1981a); Kaplan (1989b, 1989a, 1990); Kripke (1980).

3.2.3 Ellipsis

Ellipsis comes in a few different stripes. Most importantly, one should keep *syntactic* ellipsis separated from *utterance* ellipsis.²² The former is the incompleteness of the linguistic expression uttered, while the latter is an elliptical utterance of a non-elliptical expression. Thus ‘Smith went to Paris, Jones ~~went~~ to Brussels’ is clearly syntactic. But what about ‘elliptical’ answers to questions? Schiffer (1981: 69), for example, claims that in uttering ‘No’ in reply to ‘Is Harry Truman alive?’ one implicitly refers to Harry Truman. And ‘No’ is hardly syntactically incomplete. Also, suppose Keith asks Wendy ‘What is Bill doing?’ and she replies,

(26) Writing a paper.

‘Bill’ is no part of the sentence Wendy uttered although she did mean to refer to Bill rather than, say, herself. Thus Bill is a constituent of the proposition Wendy expressed.²³ However, linguists have argued that (26) *does* have a *covert* NP as subject, represented by ‘PRO,’ which is indexical or somehow anaphoric on ‘Bill’ in Keith’s question (cf. Haegeman 1994: ch. 5). If this is true (26) is not a case of ellipsis at all. But this would only mean that (26) underspecifies content in a manner similar to indexicals—the expression representing Bill just happens to be aphonic.

The more interesting case, then, is utterance ellipsis. Quantificational NPs provide an example.

(27) Everyone is going.

Wendy might utter (27) to tell Bill that everyone *in the department* is going *to the party* (cf. Neale 1990: 95). But the meaning encoded by the lexical item ‘everyone’ is insufficient to constitutively determine what particular domain of people the speaker intends in uttering (27). Failure to restrict what is said in uttering (27) to some sub-domain of *everyone* would, seemingly, have the speaker saying that everyone in the entire universe is going. Which is absurd. Thus, although ‘the department’ is not a component of what Wendy utters *the department* is, in some

²²Cf. Neale (2004: 98–105); Ostertag (1998: 20, 1999: 144n3).

²³Adopting some helpful terminology from Perry (1986), propositions have ‘constituents’ (properties, individuals, relations, etc.) and these can correspond to specific ‘components’ (predicates, names, verbs, etc.) of the expression uttered.

sense, a constituent of what she said, so, there is a part of what she said that corresponds to no part of that with which it was said. This clearly gives rise to underspecification.²⁴

Cases like this one and, in particular, Perry's (1986) example of 'It's raining,' have encouraged many different theorists to postulate *unarticulated* constituents

Consider also the example of genitive case constructions in English.

(28) I read John's book.

Suppose that the in uttering (28) the speaker intends to refer to the relation between authors and their works, call it R_{54} , as opposed to, say, another relation between owners and their books. There seems to be no reasonable way of construing 'linguistically encoded meaning' such that R_{54} is part of the meaning of (28) as a sentence of English (cf. Recanati 2004: 63, 2010: 183).

Finally, reference to times or time-spans also gives a nice example of ellipsis.

(29) I've had breakfast.²⁵

Imagine that Wendy utters (29) in response to Bill, who asked 'Do you want something to eat?' Now, one would presume, what Wendy is *really* saying might be represented by an utterance of

(30) I've had breakfast today.

Still, the temporal specification of (30) is not a component of (29). But it would be highly unusual to utter (29) and mean what someone uttering (31) would usually mean:

(31) I've had breakfast at least once in my life.

²⁴Thus *the department* is, in Perry's (1986) sense, an *unarticulated* constituent of the proposition expressed by an utterance of (27). Many theorists have taken this line and postulated unarticulated constituents to explain cases of underspecification (e.g. Neale 2007; Recanati 2002, 2010: ch. 3) while others have opted for a more syntactically oriented approach where any constituent must correspond to something in the sentence, if not at surface form then at least at logical form (e.g. Stanley 2007). I will sidestep these issues here. It is important to see, however, that unarticulation is only one type of underspecification among many. In particular, the underspecification of illocutionary force is not really a type of unarticulation (cf. §3.2.4). Thus there is no hope for describing all types of underspecification in terms of unarticulated constituency. I will say more about this in §5.3 below.

²⁵Sperber & Wilson (1986/1995: 189–190). Cf. Bach (2005: 35–39); Recanati (1993: 260, 2004: 87–90).

Be that as it may, (29) could of course be uttered to express something closer to the content of (31) than it would be to the content of (30).²⁶ If we were to somehow incorporate ‘today’ in (30), on account of its immediate obviousness, into the linguistic meaning of (29) the latter would not lend itself to the saying of what would usually be expressed by (31). An example where (29) is closer to (31) than (30): Bill actually uttered ‘I bet you’ve never ever had breakfast,’ implying that Wendy has too much trouble waking up in the morning. I conclude from this that the linguistic meaning of (29) is neutral as to whether it will have the content of (30) or (31). The elliptical nature of (29) ultimately consists in this neutrality, for if indeed (29) could *only* have the content of (30) the additional articulation of ‘today’ would become superfluous.

Usually, however, some specific time-span will be meant by Wendy when she utters (29). This time-span is underspecified by (29) considered as a sentence type, but the context can be so absolutely clear as to the relevant timeframe that actually specifying it, as in (30), would be to break one of Grice’s conversational maxims, namely that one’s contribution should not be more informative than is required (1975: 45). And if this Gricean point is allowed, the addition of ‘today’ could mean that the speaker intends the hearer to infer that breakfast would be much appreciated *tomorrow*. Since the time of utterance is commonly assumed to be known by all parties involved in a conversation, it is not surprising that articulating the time in the utterance itself may trigger conversational implicatures of this kind.

3.2.4 Illocutionary force

The linguistic meaning of a sentence σ underspecifies the mood or illocutionary force with which a speaker utters σ on a given occasion. The force, or ‘function’ more generally, may be *literal, metaphorical, ironic*; one may be expressing a *request, prediction, question*; one may also just be exercising for a play or song. These radically different forces or functions are compatible, it seems, with there being no distinguishing factor at the level of the meaning of the expression type or even at the level of prosody or intonation (Gibbs & Colston 2012: 186). Thus ‘You’ll be there’ can easily have the force of a request, prediction or a question on different occasions of utterance.

²⁶‘The content of x ’ is shorthand for ‘the content expressed by uttering x on a given occasion.’

As indicated above, this phenomenon can be used to give a very general argument for (T). The argument is a simple *reductio*. Assume that there really is a linguistic convention or regularity such that the meaning of a given expression can constitutively determine illocutionary force. Such a convention gives rise to a vicious regress and the assumption must be rejected.

Take irony. Assume, first, that there is a regularity in the speech community such that speakers may utter (32) ironically.

(32) He's a fine friend,

So, by uttering (32) on a given occasion the speaker may intend to communicate *that he's a terrible friend* (roughly).²⁷

Now, let's assume further that the convention or regularity in question is quite explicit in the language itself. In this language *L* speakers will always prefix a trope markers to their sentences. Thus, for instance, instead of 'He's a fine friend' one would utter (33) to be clear that one is *not* being ironical.

(33) I mean the following literally: He's a fine friend.

Speakers of *L* may utter (33) to make sure no irony at all is suggested. But, clearly, this kind of prefix can't carry any such weight by itself. If *L* is a natural language it is likely that (33) itself could be uttered non-literally. Is the trope marker in (33) also meant literally? And if so, will the *L*-speaker not need a metatropical metamarker to make sure this is communicated? If so we appear to have an infinite regress. An 'irony marker' would obviously have the same consequences.²⁸ Since such an explicit convention cannot fully determine the force of

²⁷The example is from Grice (1989: 53–54). Since saying is always intentional on the Gricean view, and a speaker who utters (32) ironically does not *mean* that anyone is really a fine friend, Grice suggests that that's only something the speaker *makes as if to say* or *pretends* to say. What the speaker really *said* is the non-literal, ironic content. This clearly fits my general argument in the dissertation and I fully endorse this part of the pretence theory of irony—and possibly other tropes as well. However, we need not accept this relatively controversial claim to make the present argument. But see, in particular, Camp (2012) and Recanati (2004: 18–19).

²⁸A similar point is made by Neale (2004: 93n27). For an early discussion of this type of underspecification, see Sperber & Wilson (1986/1995: 180). But the thought certainly originated with earlier authors, especially Austin and Wittgenstein. Here is Austin for example: "... there is no kind of sentence which as such is surprising, or doubtful, or certain, or incorrigible, or true" (1962: 111).

an utterance I take it that any *tacit* convention, depending on emphasis or tone of voice for example, cannot do so either. Such a convention can always be successfully violated. If Gibbs & Colston (*ibid.*, 53) are to be trusted, empirical work definitely supports this conclusion: “To date [...] no reliable procedure has been established for irony identification.” Importantly, however, ‘identification’ an epistemic notion here and not a constitutive one, so the support is only indirect, by way of the *audience*-directedness of communicative intentions.

Underspecification of illocutionary force is ubiquitous in natural language. But I do not wish to deny the possibility of cases in which the linguistic meaning of a sentence *coincides* perfectly with what the speaker said. Thus, in many cases, there may be no better reply to the question ‘What do you mean?’ than saying that one means exactly *what the sentence itself* means. Perhaps mathematical sentences provide an example of this. However, as Kent Bach (2005: 26–27) makes clear, even if such a coincidence occurs *it is not determined by the linguistic meaning of the sentence* but by the intentions of the speaker. The fact that (33) is uttered literally as opposed to ironically must be explained by something other than only the meaning of (33) as an expression type. If this phenomenon is indeed ubiquitous one must conclude that linguistic meaning essentially or always underspecifies propositional content. So (T) is true.

This is where Gricean intention-based semantics can step in and offer a plausible theoretical explanation. The encoded meaning of a linguistic expressions is abstract and informationally impoverished, only placing constraints on the kinds of communicative intentions the competent speaker can form in uttering it on a given occasion. It is a blueprint or template embodying instructions for constructing full-blooded propositional contents. As Grice argued, the constraints are determined by regularities of use within a community: if expression σ has meaning m in language L then speakers of L have a “procedure in their repertoire” to utter/hear σ intending/understanding it as having meaning m (Grice 1989: 127–128; Schiffer 1972: ch. 5). But that doesn’t preclude other procedures where σ has another meaning, loosely related or entirely different.²⁹

²⁹This line of thought is continued in Chapter 5 below, where the notion of a propositional blueprint is fleshed out in more detail.

3.2.5 What about so-called ‘eternal’ sentences?

One important objection to (T) should be considered before moving on. According to this objection, which can be traced back to Quine (1960: 193–94) and Katz (1972), there are so-called ‘eternal’ sentences and the only reason (T) seems right is because most sentences actually uttered are in fact mere *abbreviations* of such sentences. Arguably, then, such underspecification as can be found in natural language is to be explained away as a simple matter of effort-saving convenience. Whenever the meaning of σ underspecifies the proposition *that p*, there exists another corresponding sentence σ' whose meaning fully determines p , such that the speaker *could have* uttered σ' instead of σ in the context in question.

Let’s state the Quine-Katz argument more clearly. First, define ‘eternalization.’

Eternalization

Sentence σ in language L eternalizes utterance U iff the proposition p expressed by the speaker in making U is such that any utterance of σ must also express p .

Note that this defines a relation between dated *utterances* and sentence types. To illustrate, assume that the speaker utters

(34) He went to the bank.

Thereby expressing, say, the proposition *that Simon Kirchner went to the financial institution at 2pm on 29 May 2014*. Call this message or proposition p for short. Then, it seems safe to assume, (35) gives a possible eternalization in English.³⁰

³⁰Mathematical sentences have also been used as examples of eternal sentences. The sentence,

(1) $2 + 2 = 4$

appears to express the same proposition whenever it is uttered. But I don’t think (1) is a good counterexample to (T). For one thing, there is a huge controversy about what mathematical expressions actually mean. Do they refer to numerical objects? Do they express rules? Are they truth-apt? I will not pretend to have answers, but I would say that mathematical expressions are formal rather than natural. Thus they are stipulated always to mean the same thing. Stipulations are conventions, so we should mostly encounter abbreviation rather than underspecification. Furthermore, if (1) can be uttered to express propositions—with numbers as constituents—then it is not immune to underspecification in terms of force or function. (1) can be a question or a command (to write down (1) say). And if numerals can be uttered to refer to numbers of different *kinds* (real numbers, natural numbers, etc.), then this gives rise to underspecification as well.

(35) Simon Kirchner went to the financial institution at 2pm on 29 May 2014.

And surely p is not underspecified by (35), is it? We should certainly acknowledge the plausibility of (35) being eternal, i.e. that on all occasions on which a speaker utters (35) she will have expressed p . (Assume for now that only one person has ever been named ‘Simon Kirchner.’)

But does the existence of (35) disprove (T)? I don’t think so. Robyn Carston (2002: ch. 1), drawing on arguments in Wettstein (1979), has argued that eternalizations or are not so easy to generate as many, including Quine and Katz, have assumed. Thus, for example, when a speaker utters (34) on a particular occasion there just is no single eternalizing completion like (35) which accurately captures the temporal constituent of the proposition expressed.³¹ Did the speaker intend 14:00 hours as opposed to 2pm? Or 14:00:00 hours as opposed to 14:00:01 hours, the first being the time at which the utterance started and the latter when it finished? Most likely the speaker meant none of the above and no precise eternalization is in the offing. Carston develops this thought at length to argue, basically, for her version of thesis (T). Although I clearly agree with the conclusion, I think it can be reached *without* begging any questions about the possibility of eternalizing dated utterances. Plausibly, then, we can accept that this is an epistemological problem: yes, it’s very hard to formulate or discover the exact eternalization, but it still exists in some sense—it’s something to which our inadequate forms of expressions are pitiful approximations.

Finally, the Quine-Katz argument can be semi-formalized like this.

P1 Most (or maybe all) actual utterances of sentences in a given natural language are such that the sentence underspecifies the proposition thereby expressed.

P2 However, for every utterance U there exists a sentence type σ that eternalizes U .

C3 Thus underspecification is merely a matter of effort-saving convenience.

Carston argues that P2 is false and the argument unsound. But the argument is in fact invalid and C3 doesn’t follow even if P2 is conceded. To see this, consider a small thought experiment. In a way similar to Borges’ well-known story, ‘The Library of Babel,’ assume there exists an

³¹See Buchanan (2010) for a recent formulation of roughly this type of argument.

enormous *library of thoughts* in which every possible proposition is listed and assigned its unique proper name (e.g. by some number system). Assume further that humans are endowed with infallible memory and flawless intelligence. Thus we inhabit a world in which humans have recourse to eternal signs for all propositions, which their actual utterances, in some sense, *abbreviate*. But for some reason—brevity? fashion? tradition?—speakers still utter natural language sentences just as before. Clearly, they *could always* supply another sign which eternally encodes or refers to what they said. It does not follow, however, that the sentences actually uttered don't underspecify the propositions thereby expressed. The mere fact that there exists some other uniquely determinative form of expression changes nothing in this respect.

There is a world of difference between abbreviation and underspecification. The former is usually a relation between two things of the same type, e.g. two expressions. Thus the expression 'CEO' abbreviates the expression 'chief executive officer.' Such relations are symmetric, so that the latter expression will actually 'lengthen' the former. Such a relation is also *conventional* or, even, on some accounts, *analytical*. It is a linguistic *convention* of English, if anything is, that 'CEO' abbreviates 'chief executive officer.' And if Russell's theory of descriptions is right sentences of the form 'the *F* is *G*' abbreviate—i.e. are analyzed as—conjunctions of three sentences which can be stated more succinctly in first order logic: $\exists x(Fx \ \& \ \forall y(Fy \rightarrow x = y) \ \& \ Gx)$. But underspecification is most naturally construed as a relation between a sentence and *what is said* by a speaker who utters it on a given occasion. On some accounts at least, one is a linguistic entity and the other a proposition.

Putting ontology aside, however, it is not reasonable to think that sentences and propositions stand in any conventional relation akin to abbreviation. First, there can be no linguistic convention to the effect that (34) expresses *p*, for (34) can be uttered to express an infinite number of different propositions on different occasions, and it would be futile to suggest there exists a unique convention for every occasion. If there is a convention in the vicinity, it will only determine *that* there must be a specific time at which Simon went to the bank, not *which* particular time is referred to on a given occasion. Second, the sentence uttered does not abbreviate its eternalization, even if they are at our disposal (e.g. in a Borgesian library). True, when a speaker utters (34) she may communicate something that she could *also* have com-

municated, in the same context, by uttering (35). This, however, is a far cry from saying that one is a conventional abbreviation for the other. Even if by the words ‘It’s raining’ a speaker may mean *that it’s raining in Paris*, those words—type or token—are not conventionally short for ‘It’s raining in Paris.’³² And the latter is not a conventionalized ‘lengthening’ of the former either, which it would have to be if underspecification were akin to abbreviation.

What is worse, the Quine-Katz argument seems to require three substantive auxiliary assumptions to go through, even if other problems were to be ignored. One, that there is a language of thought, the medium of which is, at least partly, some *public* language.³³ Two, that sentences in our language of thought are eternal. Three, that thoughts or beliefs are constituted by mental tokens of such sentences. I’m happy to concede the first two, but the third is problematic.

Briefly, if this assumption were accepted, we’d have a robust relation between sentences uttered and the propositions they abbreviate. For example, a speaker who utters something like (34) would not express anything like *p* but would, rather, *express the sentence in* (35)—given, again, that (35) actually eternalizes the target utterance relative to the speaker’s idiolect and the occasion in question. Instead of P2 in the Quine-Katz argument, we’d get something like:

P2’ For every utterance *U* there exists a sentence type σ that eternalizes *U* and the speaker expresses σ by making *U*.

However, speakers do not utter short expression in order to convey or indicate some longer expression. Speakers do not have any intention, with respect to an expression like ‘CEO’ that it should express ‘chief executive officer’ on a specific occasion. Both would normally be intended to refer to a particular person. Speakers do, however, regularly utter the shorter expression to say something that could just as well have been said by uttering the longer expression.³⁴

³²I discuss this issue at more length in Unnsteinsson (2014: §6.2).

³³Fodor (1975, 2008); Harman (1973). See Devitt & Sterelny (1999: 138–146) on the issue of public language versus Mentalese as the medium of thought.

³⁴François Recanati (2010: 124) argues that, in order to avoid violating what he calls ‘the principle of full articulation,’ token expressions must take on specific ‘occasion meanings’ in context. Thus a token of ‘eat’ can mean ‘eat dinner’ and a token of ‘drink’ can mean ‘drink alcohol.’ The criticism developed here applies directly to his theory as well, cf. Unnsteinsson (2014: 3346–3347).

More to the point, something like the third assumption receives its most worked-out defense, to my knowledge, in work by Peter Carruthers (1996) where he pits together ‘communicative’ and ‘cognitive’ conceptions of language, with a deep-rooted preference for the latter. But even in Carruthers’ final analysis the theory is far too weak to sustain P2’ or anything of the sort. On the cognitive conception, natural language is the *primary* medium of *conscious* thought. So there are other mediums and there are other thoughts, so, the view is only that language is constitutively *involved* in some thoughts and beliefs (see pp. 2, 51, 120–123). But this is acceptable to both sides and not sufficient for P2’. Obviously so, for otherwise eternal sentences would need to be the primary medium of our conscious thoughts which, it seems, is just absurd.

In closing, remember that if the argument from underspecification is correct, infinite regress makes the eternalization of, at least, illocutionary force impossible. Carston (2002: 30) argues that we can progressively make sentences ever closer to determining a unique content so they all but embody the propositions expressed, but this approximation will essentially be asymptotic. Introducing a new element into the sentence, especially elements intended to determine literalness, irony, metaphor, and the like, will only invite new ways in which the proposition expressed becomes vulnerable to underspecification.

I take it, then, that (T) should be accepted and that this indirectly supports the intentionalist view of meaning and communication, since it construes sentence-meanings essentially as evidence for, and constraints on, communicative intentions, which actually determine what is said by the speaker.

3.3 Malapropism: Saying without meaning?

Malapropisms are commonly used as counterexamples to intentionalism. We defined intentionalism, roughly, as the view that when speaker *S* says *that p* by uttering sentence σ in language *L* on some occasion, (i) *p* must be constitutively determined by *S*’s speaker meaning on that occasion and, (ii) *p* must be compatible with the abstract linguistic meaning of σ in *L*. Speech errors of various kinds have been taken, by many, to show that speakers actually say *that p* without speaker meaning *that p*.

In this section, I explain how this kind of argument works. The most fully worked out case of this kind is due to Marga Reimer. She argues that her own ‘conventionalist’ theory of saying, and of what is said by speakers who utter malapropisms, gives the best explanation of the linguistic data. Then I present an objection to her theory, based on the argument from underspecification. I then present my own theory, the misarticulation theory of malapropisms, and show how it is supported by work done in this area by phoneticians. Finally, this makes it possible to state another objection to Reimer’s view; the argument from arbitrariness.

It is difficult to do justice to the variety of slips and verbal blunders falling under the heading of ‘malaprop,’ so, to fix ideas, I focus only on three kinds. Note that the structure of this objection to intentionalism mirrors perfectly the confusion-based arguments already debunked in Chapter 2. The difference is, most importantly, that malaprop-based arguments need not involve speakers with false identity beliefs. But they do require speakers who unintentionally perform some sort of speech error.

3.3.1 Objection to intention-based semantics

Consider the following cases.

1. *Incidental malaprop*: When John Kerry was presidential candidate for Democrats in 2004, running against George W. Bush, he slipped while giving a speech and uttered ‘wasabi’ instead of ‘Wahhabi.’ Kerry clearly intended to refer to a Muslim fundamentalist sect called ‘Wahhabi’ but, due to fatigue or whatever, the similar-sounding ‘wasabi’ stumbled through.
2. *Persistent malaprop*: Reimer (2004) describes a colleague who persistently uttered ‘obtuse’ when he clearly meant that something was *abstruse*.
3. *Intentional malaprop*. Davidson (1986) cites an example where the speaker/writer intentionally utters/writes ‘baffle of wits’ instead of ‘battle of wits’ for comic effect.

Any of these examples, and others, can then be used to make a plausible argument, along the following lines.

For the first two types there is some proposition p such that the speaker actually said and asserted *that p*, without meaning *that p*. In the first example, Kerry said something about *wasabi* without meaning anything about *wasabi*. In the second the colleague said someone's writing was *obtuse* without meaning that anyone's writing was *obtuse*. Describing the third type of case is somewhat more complicated, so I leave it until later. But it follows straightforwardly from this description of the other two cases that speakers say and assert things they don't intend or mean to say or assert.

Question: Why should this description be taken at face value? Citing the 'obtuse'-example, Reimer (ibid., 322) writes that "[t]he speaker of a malaprop, upon being informed of the fact that his use was non-standard, would likely agree that what he actually said was different from what he intended to say." A few sentences later, she gives a fairly direct answer to my question.

[I]f we are going to develop a philosophically sound notion of saying, it would presumably be best to build it upon a pre-theoretical notion that is sensitive to a distinction that has clear explanatory value: the distinction between saying and (speaker) meaning. This is a distinction that allows us to explain (*inter alia*) the coherence of claiming that one doesn't always mean what one says.

According to Reimer, then, we should posit a distinction between saying and meaning to explain why it is coherent for speakers to describe malapropisms in terms of the distinction.³⁵ But how is what is said determined, then, if not by way of communicative intentions? On Reimer's view, it is determined by *linguistic convention*; in making the malaprop the speaker simply *says* and *asserts* what the words 'conventionally mean' in the language in question. Without assuming a full-blown theory of conventionality, Reimer takes this to imply, at a minimum, that by engaging in the rule governed activity of speaking English, the speaker tacitly agrees to have his utterances interpreted in accordance with the conventions of the language. As they say, speaking a language is like playing a game. It follows, then, that Reimer's colleague asserted that someone's writing was obtuse while meaning that it was abstruse.

³⁵Michael Devitt (2013d: 88) also describes speech errors, spoonerisms in particular, as cases of unintentionally saying one thing and meaning another.

More specifically, Reimer—and those who share her general outlook³⁶—speaks of ‘contextually relativized conventional meaning’ in order to “accommodate indexicality and ambiguity” (ibid., 333n2). Assuming that this is clear enough let’s the Reimerian thesis as follows:

Conventionalist theory of malapropisms

When speaker *S* utters malaprop-sentence μ , the content of what *S* says is determined by the contextually relativized conventional meaning of μ , which is, say, the proposition *that p*.

Additionally, *S* may mean and implicate all sorts of things other than what *S* strictly says by uttering μ . Further, at least when dealing with *unintended* malaprops, *that p* is *no part* of what *S* means or intends.

The theory fails, but it is important to see why exactly this is so. In light of the prior discussion of underspecification, let’s take a closer look at so-called ‘incidental’ malaprops. Assume Kerry uttered,

(36) Wasabi is a dangerous sect,

but his plan was to utter,

(37) Wahhabi is a dangerous sect.

Now, what exactly, according to the conventionalist, did Kerry *say* by uttering (36)? What is the contextually relativized conventional meaning of (36)? The simple disquotational answer would be that Kerry said (38).

(38) *that wasabi is a dangerous sect.*

But this is far from clear. ‘Wasabi’ is clearly polysemous or ambiguous.³⁷ It can refer to (i) a condiment popular on sushi, or (ii) a plant of the *Brassicaceae* family, from which the condiment is produced. The conventionalist has three options, it seems: either Kerry said (i), or (ii), or (iii) the meaning is somehow indeterminate between the two. But all three options are problematic.

³⁶This group is of course large and heterogeneous. I take Searle (1969: ch. 2) to be a classic statement of this type of conventionalism. But see also note 15 in Chapter 2.

³⁷The difference between polysemy and ambiguity seems to be mostly a matter of degree. See Sennet (2011) for discussion.

Assume first, and quite plausibly, that in the past Kerry has usually uttered ‘wasabi’ to mean (i) and is only vaguely aware that ‘wasabi’ is also the name of the plant. Let’s now go through the options one by one. First, is it possible that Kerry said something about the condiment by uttering (36)? It’s glaringly obvious that *relativizing to context* doesn’t support this answer. *Ex hypothesi*, there is nothing at all in the immediate context to determine (i) as opposed to (ii). The utterance is made in the context of a speech that has nothing to do with any plant or condiment.

What if the contextual salience of (i) is increased by supposing, e.g., that Kerry actually said *that he didn’t like the taste of wasabi* right before he uttered (36), making the same error? (Inventing a plausible context is surely possible here.) This would still not suffice to determine which meaning is at issue. Polysemous expressions can quite easily be intended in different ways in one and the same context. So the following conversation—where the speakers refer successively to the condiment and the plant—shouldn’t raise any eyebrows.

(39) Wasabi is so expensive.

(40) Yeah, wasabi is very hard to cultivate.

Someone might stop me here and say, “You’re already assuming that *wasabi-as-condiment* is much more salient in Kerry’s mind and in the context of utterance. So, arguably, only this meaning would play a role in any *causal explanation* of Kerry’s malaprop. If so, wouldn’t this suffice to determine the contextually relativized conventional meaning of ‘wasabi’ in (36)?” Sounds plausible to me, but note, however, that taking this line is not open to the conventionalist. Conventionalism, as it is understood here, is predicated on minimizing the role of causal chains in determining contents. Remember that, on Reimer’s account, language users tacitly agree to have their words interpreted in accordance with prevailing conventions, relative to context. And letting causal history override convention doesn’t leave much over for convention, explanatorily speaking (more on this on p. 98 below).

The same set of considerations counts against option (ii) and more strongly so, as it’s the less salient of the two meanings. What about (iii)? Is the conventional meaning of the malaprop indeterminate between, at least, (i) and (ii)? Whatever we say about the case at hand, this cannot be the conventionalist’s answer in general. Consider a starker example where the

speaker planned to utter ‘Take those dead bats to the tank’ but, in the psycholinguistics jargon, performs the speech error of ‘perseveration’ and utters (41) instead.

(41) Take those dead bats to the bank.

Perseveration is when an earlier segment of speech replaces a later item, so, here, /b/ replaces /t/ (cf. Carroll 2007: ch. 8; Fromkin 1973; Levelt 1993: ch. 9).³⁸ Now, if conventional meaning always underspecifies the meaning intended by the speaker, the sound sequence /bank/ is quite clearly ambiguous and can be intended, by normal speakers on different occasions, as referring to a financial institution or a riverbank. As before, the context is assumed to have nothing to do with either conventional meaning, so it can’t help the conventionalist. But if option (iii) is endorsed, the theory predicts that what is said by (41), relative to some appropriate context, is indeterminate between the two meanings. This consequence, when generalized, is clearly unacceptable to the conventionalist. Why? Because, for a lot of cases, the theory would be left with no single, sensible answer to the question: What is said by the speaker of the malaprop-sentence?

The problem here, of course, is that the conventionalist lacks a trick the intentionalist has up his sleeve. On the intentionalist view, sentence-meanings place constraints on the kind of communicative intention a speaker can form in uttering a sentence with that meaning on some occasion. But what is said by the speaker on an occasion is constituted only by the communicative intention, although the intention needs to be compatible with the meaning of the expression the speaker *planned* to utter. With this background one can always fall back on saying, even when the speaker is being indeterminate or engaging in *double entendre* of some sort, that the speaker’s intention determines which meaning is at play. True, this means the intentionalist is saddled with the counterintuitive thesis that there is no saying without meaning. If the arguments in this dissertation are on the right track, however, this is much less of a problem than many theorists have tended to suppose.

³⁸Because of the idiomatic expression ‘can take it to the bank’—meaning, roughly, that someone can depend on the truth of what one says—the slip in (41) might also be explained as a case of ‘substitution,’ where part of the sentence uttered is infiltrated by some unrelated expression.

3.3.2 The misarticulation theory of malapropisms

Now I present what the Gricean theorist, in a more positive way, should say about malapropism. The idea is basically to use the model of *mispronunciation* on all varieties of speech error. This results in a parsimonious and elegant theory that is compatible with the full range of ‘speech errors’ revealed both by experimental work and empirical observation.

Misarticulation theory of malapropisms

When speaker *S* utters malaprop-sentence μ , the content of what *S* says is determined by *S*’s communicative intention in uttering μ , for example *that p*. This is so even if μ happens to be standardly uttered, in *S*’s language, to say something other than *p*.

So even if it is normal, in *S*’s language, to utter μ with some different communicative intention, which ‘fits’ the conventional meaning of μ better, it doesn’t follow that that’s what μ says on a given occasion. Further, when *S* misspeaks there is always some expression σ in *S*’s language such that *S* misarticulated σ as μ and the proposition expressed by *S* in uttering μ need only be compatible with the abstract linguistic meaning of σ , on the occasion of utterance. Indeed, σ is the target of the speaker’s intention and, on this view, the communicative intention need only be compatible with the target expression, not the erroneous expression. One clear advantage of this theory is that it captures the full range of speech errors, also those made by, for example, people learning a second language. Native speakers of Icelandic, when learning English, tend to confuse /v/ and /w/ in speech articulation. So my wife sometimes says she is in ‘Vest Willage’ rather than West Village. Hearers have no problem compensating and can easily understand what is being said.

More generally, empirical work on speech perception reveals what linguists call the ‘lack of invariance problem,’ namely that there is no one-to-one correspondence between acoustic signals and perceptual categorization into phonetic segments (e.g. Appelbaum 1996). Speech perception thus involves all sorts of automatic mechanisms compensating for context-induced variation in acoustic cues. Philosophers have tended to overlook malapropisms where such compensation is automatically or unconsciously performed by the hearer and therefore, arguably, their perspective has been skewed towards cases where speakers may appear to say

something other than they intended. Furthermore, experimental work on misspeaking shows a tendency for speech errors to produce words rather than non-word strings. This is called the lexical bias effect (see Levelt 1993: §9.5.2. for discussion).³⁹ My claim is that malapropism should be seen in light of articulatory variance more generally.

So, to keep to the same example, the native speaker of Icelandic might utter, when arriving at a dinner-party,

(42) I brought the vine,

and thereby mean that she brought the *wine*.⁴⁰ But the suggestion on the table is that, for the purposes of semantics, this utterance of (42) should not, merely because ‘vine’ happens to be a word in English, be considered any different from misarticulations resulting in non-word strings. The misarticulation theory claims that the speaker of (42) happened, on this occasion, to pronounce ‘wine’ as ‘vine.’ Here, and, it would seem, for the vast majority of misspeakings, this description is perfectly fine. But it should be expanded to capture the less intuitive cases as well, paying handsome dividends in terms of simplifying the overall theory.

Admittedly, cases of this sort are intuitively understood in terms of simple differences in idiolect. Any ‘public language’ is just what a bunch of idiolects share and these vary in all sorts of ways. This does not, however, speak against the present point. In so far as a speaker of (42) has /wine/ as her *target* sound sequence on a given occasion of utterance, she performs a speech error. Of course there will be cases where we can only describe the speaker as having an idiolectal /vine/ sequence as target. But think of cases where she is being ‘corrected’ by native speakers and doesn’t get it right until after a few attempts: here there is clearly a target-error pair. There is a continuum of cases going from idiolectal variation, through persistent malaprops, until we get to the one-off, incidental cases.

³⁹My two year old can only pronounce the name ‘Oliver’ as ‘olive oil’—the English is very similar to the Icelandic—which, I presume, is partly explained by lexical bias since she learned the latter first. I could give a very long list of such examples.

⁴⁰Karen Green (2001: 242) also gives a good example of a persistent malapropism that fits this theory well. In a school project, her young son wrote, “Hitler was the leader of the nasty party.” In this case the speaker misarticulates ‘Nazi’ as ‘nasty.’ Admittedly, the case is more complex, since he may also have had the (true) belief *that Hitler was the leader of the nasty party*. But if this is a malaprop at all, and not just a plain expression of this true belief, we must also assume that /nasty/ is his erroneous way of articulating ‘Nazi.’

Slips of the tongue where one proper name replaces another are perhaps the least intuitive cases for the misarticulation theory. For it predicts that a speaker might pronounce, e.g., ‘Frank’ as ‘Joanne,’ which seems absurd. But let’s start with a real-life example and take it from there. When Osama Bin Laden was killed numerous media outlets reported something like (43).

(43) Obama was killed.

It was clear that they were making a mistake. But even if many couldn’t resist giving the Freudian explanation that (43) manifested the speaker’s repressed desire, phonetic considerations are much more powerful.⁴¹ In this case the target expression, ‘Osama,’ shares a lot of features with the error expression, including stress-pattern and syllable number. Importantly, this is true of the majority of actual target-error pairs. The names also share many general semantic features, e.g. they are both names of powerful men who are often in the news. Further, there may even be a phonetic anticipation effect hiding here since ‘Osama’ can easily activate the expression ‘Bin Laden’ and this could make the activated /b/ segment replace the earlier /s/ segment.⁴² So, the speaker of (43) simply pronounced ‘Osama’ as ‘Obama.’ Certainly there will be slips where the phonetic explanation is not as obvious, but according to the empirical literature these are much less common and, arguably, they should not dictate theoretical choices. Our theory should generalize from the most common type of case, i.e. ones where the explanation is phonetic, and treat the strained cases as anomalies. This clearly applies to examples like the one where a speaker intends to utter ‘Frank’ but utters ‘Joanne’ instead.

3.3.3 Argument from arbitrariness

Accepting level-headed, phonetic explanations gives the upper hand to the misarticulation theory since the semantics of malaprop-expressions usually figure only minimally in the best explanations of their occurrence. If the speaker had uttered ‘Ofama’ instead of ‘Obama’ no one—and that includes the conventionalist—would have thought that she had unintentionally

⁴¹Considering the fact that even Ted Kennedy confused the names, the Freudian account seems far-fetched.

⁴²‘Anticipation’ is the opposite of perseveration, where a later segment replaces an earlier one, e.g. ‘bake my bike’ replaces ‘take my bike’ (Carroll 2007: 195). Michael Erard (2008: 264–267) discusses the Obama/Osama example in a book popularizing phonetic explanations of malapropisms.

said anything about anyone. But since the speaker *could have* uttered (43) in the same context intending to refer to Obama we tend to think of the case differently. Wrongly so.

Remember also our conclusion from the Wahhabi/wasabi example. If the conventional meaning of 'Obama' underspecifies its intended reference on a given occasion of utterance, which must be right I think, there is nothing to fix or determine *which* Obama is referred to in (43). According to the Gricean, only referential intentions can perform this task but, *ex hypothesi*, the speaker had no intention to refer to anyone called Obama in uttering (43). Thus the conventionalist must collapse the referential content of 'Obama' in (43) with its descriptive content, which is, presumably, something like *the person called 'Obama.'* Admittedly, in this particular case, one referent is extremely salient, but still, there is nothing in the conventionalist story to determine whether (43) is about the 44th US President or, say, Barack Obama Sr., his father.

To be clear, however, the issue is not decided conclusively by discovering the right causally explanatory theory of malaprops. Conventionalism is strictly speaking compatible with any such theory since it's only a claim about the *meaning* or *content* of malaprops in context. But empirical findings yield grounds on which to build the decisive argument, which is an argument from *arbitrariness*: the conventionalist theory makes an entirely arbitrary distinction between word and non-word producing errors. The first kind, it is thought, *says something* but the latter *says nothing* since the resulting string has no conventional meaning at all. Clearly, this content error/empty error distinction crosscuts any distinction made by a plausible explanatory theory. On anyone's account there will be some content errors that are explained in exactly the same way as any kind of empty error, namely by way of phonetic/phonological effects. Content errors are common because of the lexical bias effect. But this effect is mostly due to formal similarities in target-error pairs, such as morphological and syntactic distribution, stress-patterns, and the like. If the conventional meaning of the error expression does not play a role in explaining its occurrence, it is hard to see why it ought to be postulated as the content expressed by the speaker in uttering the malaprop-sentence on a given occasion.

Still, it might be claimed—this was indeed suggested to me by Michael Devitt—that in those cases where semantic features play a substantial role in the correct causal explanation of a

malaprop the corresponding meaning could be attributed to the speaker.⁴³ I think this is fine as far as it goes, but it doesn't go very far. First, as already noted, Reimer's conventionalism doesn't allow for this type of response. Secondly, and more to the present point, what causes the utterance of an expression cannot in general be identified with what is meant or said by that expression. Obviously phonetic features are not assigned as the contents of what a speaker says because they causally explain the fact that the utterance was produced on a particular occasion. Only the communicative intention of the speaker can constitutively fix or determine what is said and meant on a given occasion of utterance.

Thirdly, and finally, even if the causal-conventionalist thesis were accepted it would leave out a host of cases where speakers indeed uttered a malaprop which happened to be syntactically well-formed, but where the correct explanation is entirely or partly phonetic/phonological. Thus the objection from arbitrary distinctions crops up again: Why suppose that some syntactically well-formed errors are assigned contents and others are not? On the misarticulation theory, the only relevant possible world in which Kerry says (38) by uttering (36) is one where he—perhaps for one strange moment in a speech—actually believes *that wasabi is a dangerous sect*. In such a case, of course, the intentionalist assigns this very content to the speaker on the occasion of utterance. But any real-world explanation of a malaprop will be a messy affair of intermingled syntactic and phonetic features, where it would be entirely arbitrary to assign (38) to some (36)-malaprop-utterances on some occasions and not to others. Better to say Kerry mispronounced 'Wahhabi' as 'wasabi' and be done with it.

So far, nothing has been said about *intentional malaprops* like the one mentioned by Davidson (1986: 89) where the speaker utters 'baffle of wits' instead of 'battle of wits.' This is as it

⁴³See also Devitt (1981a). After discussing (combinatorily) confused reference and, plausibly, explaining them in terms of his theory of causal networks of uses grounded in the object referred to, Devitt turns to malapropisms: "[...] more than one ability, and hence more than one network, may have an immediate role in the production of a designational term. When this happens, we have slips of mind or tongue, cases of "crossed wires." A classical example of such an occurrence was supplied, appropriately enough, by Canon Spooner. He once delivered a sermon that included many uses of 'Aristotle.' He was leaving the pulpit when suddenly he stopped, returned, and announced to the congregation, "When in my sermon I said 'Aristotle' I meant St. Paul." We are inclined to say that Spooner had St. Paul in mind but ended up referring to Aristotle" (p. 139–140). Indeed we are so inclined. But the inclination ought to be resisted in this case. Spooner, for some reason, and over some specific stretch of time, mispronounced 'St. Paul' as 'Aristotle.'

should be, because these present problems of their own and need a separate treatment. They are malaprops in name only for the speaker doesn't make any mistake at all; he uses an expression the utterance of which *would* constitute a mistake in some different context, e.g. a context where the speaker falsely believes that 'battle' should be pronounced /baffle/. More importantly, for our purposes, intended malaprops don't provide occasion for disagreement between the conventionalist and the misarticulation theorist. On Reimer's account (44) gets assigned its contextually relativized conventional meaning, namely *that a baffle of wits ensued*.

(44) A baffle of wits ensued.

So 'baffle' means *baffle* and 'wits' *wits*. Certainly, Reimer adds, the speaker also communicates, by association, that what happened was a *battle of wits* and that it was in some way cause for *bafflement*.

The point to make here is that the intentionalist is free to concur, and postulate that the speaker's communicative intention in uttering (44) may very well be *that a baffle of wits ensued*. It's just that this happens also to be its conventional meaning, according to Reimer. The disagreement would occur only at the level of content determination. The conventionalist assigns this meaning *because* it is the conventional meaning. The intentionalist does so because it is the meaning intended by the speaker. This kind of case is so very special, it seems, that it cannot be used to adjudicate between these two theories.

At any rate, here is how I would describe the case. Very briefly, 'baffle of wits' in (44) expresses the *ad hoc* concept *baffling battle of wits* or just *baffle of wits**. Call this concept *C*.⁴⁴ *C* is *ad hoc* relative to the uttered expression 'baffle of wits,' but it is actually encoded linguistically by the longer expression 'baffling battle of wits.' Since it has certain positive cognitive effects—e.g. being funny—to express *C* with the shorter expression rather than the longer, it was chosen by the speaker of (44) on this occasion. But this kind of account is clearly

⁴⁴An *ad hoc* concept *C** can be defined in terms of the present state of a public language *PL*: At *t*, *PL* is such that *C** is not lexicalized and the most efficient or appropriate way to express *C** in some concrete context is to utter a simple expression that encodes a related concept *C* (such that *C** implies *C* or *C* implies *C** or ...) intending that the hearer will immediately infer that *C** is what is actually meant. See, in particular, Wilson & Sperber (2002b: 72–76) on the concept of FLATNESS. Also Barsalou (1983, 1987, 1999); Carston (2002, ch. 5); Prinz (2002, ch. 6). *Ad hoc* concepts are discussed further in §5.2.4 below.

not compulsory for the misarticulation theorist.⁴⁵

Bearing in mind Austin's distinction between doing something by mistake and doing something by accident, malapropisms fall squarely in the latter category.⁴⁶ Speakers have a specific target in mind but the articulatory mechanism fails them, either persistently or just incidentally. If the argument in this section is correct, such 'accidents' are far more common than people tend to think. Furthermore, Gricean intention-based theories of communicative content can account for them without abandoning the idea that what is said is constitutively determined by speaker meaning. It seems plausible to say that malapropisms provide cases where speakers say something they didn't intend or mean to say. But this is an explanation from folk psychology that can easily be restated in terms acceptable to Griceanism, without loss of explanatory power. When a speaker performs a speech error and produces a syntactically well formed sentence different from the one he intended, he doesn't thereby *say* what that sentence is normally taken to say. He simply produces a sentence that could have been uttered—if the speaker had had the requisite communicative intentions—to say what it is normally taken to say.

In terms of the bigger picture, the case for the misarticulation theory of malapropisms shows the acceptability of idealizing away from pragmatic performance errors when theorizing about foundational notions in semantics, such as *saying*, *meaning*, *implicating*. Same goes for the Austinian category of linguistic utterances that are made 'by mistake.' Even such cases should be described as a special kind of pragmatic error, making the semantic conclusions theorists are aiming for much less immediate.

Conclusion

Time for a breather. How is all this connected to the arguments in Chapters 1 and 2? The connection is twofold. First, malapropisms and confusion-puzzles have both been used by philosophers to argue that speakers can say things they do not mean or intend. The basic

⁴⁵For what it's worth, I think this type of account is closer to capturing Davidson's positive view in his 1986 "A nice derangement of epitaphs" than Reimer's (2004: 325) interpretation. But I won't argue the point here.

⁴⁶Cf. Austin (1957: 133n1). See p. 2 of the Introduction.

Gricean point, in response, is that there is a more parsimonious description in the vicinity, namely that speakers can utter a string of words without thereby meaning what those words would normally be taken to mean by hearers in context. Thus there might be rational, well-informed hearers who would have good reason to suppose that the speaker actually said something she in no way meant to say. And, as should be clear by now, the same point applies to ‘actual’ reference and ‘intended’ reference. At this point in the dialectic, I consider it to be established that neither malaprops nor utterances of confused speakers supply grounds for thinking that there are sayings without meaning, or acts of referring without referential intentions.

If, as suggested in Chapter 1, singular terms have the proper function of constituting evidence for the speaker’s referential intention, this function is clearly disrupted when speakers commit unintended malapropisms. This is the basic way in which intended malaprops differ from unintended ones; the latter always point to some sort of malfunction. To give a brief gloss, when a speaker *S* has a phonetic articulatory plan of uttering the sequence /osama/ but, for some reason or other, it comes out sounding like another lexical item in *S*’s language (or like none at all), the utterance will in principle provide bad evidence for *S*’s communicative intention.⁴⁷

Furthermore, philosophers’ interest in malaprops is warped by their clinging to a puzzle-driven methodology in the philosophy of language. Paralleling their reason for focusing on identity confusions, they have framed their task, here, as finding a coherent system of content-assignments to malaprop-sentences in context. This goal is certainly not worthless, but it must serve other more general theoretical tasks. The most general task is, in my view, to explain the fact of successful linguistic communication. Now, confusions and malaprops have both been shown to disrupt the evidential function of the lexical items involved. It is for this reason that these cases ought to be treated as *special*, in need of *separate* treatment: theorists should not jump to semantic conclusions based on consulting intuitions—even if they are robust and shared—about these types of cases.

Finally, finding out which kinds of mental states or speech errors have disrupting effects

⁴⁷For a detailed discussion of phonetic plans in speech production, see Levelt (1993: ch. 8).

on the evidential function of specific lexical items helps us to isolate a notion of successful communication that's metaphysically basic. By this I mean that we can formulate a notion of the act of referring—and the act of saying—that stands a chance of playing a basic role in a theory of communication. This involves some degree of idealization, or Carnapian explication, of the folk conception of referring. Of course, this has already been done, to some extent, in the Gricean tradition; the suggestion here is that the process of refinement needs to go further. But this is the topic of the next Chapter.

Chapter 4

Explicating speaker reference

Introduction

Now it has been shown, or so, at least, I will presume, that what is said by a speaker on an occasion must be fixed by the speaker's communicative intention. We have also seen, in general, how a Gricean analysis of such communicative intentions looks like. But what is the correct intention-based theory of *referential* intentions in particular? What follows is an attempt to answer this question.

First, I introduce the notion of 'optimality' as it occurs in Grice's own theory of meaning-intentions, arguing that this notion fits nicely with Millikanian analyses in terms of proper functions and normal explanations of the maintenance of such functions in a population. Secondly, I propose a Gricean explication of the mental state involved when a speaker refers to an object with a linguistic expression. Thirdly, I introduce a constraint on any notion of this sort called the 'edenic constraint.' I then show (§4.1.1-4.1.4) how the arguments in earlier Chapters provide support for the constraint, adding two more arguments in the process. I also introduce a constraint on coreference, giving one argument in its favor (§4.2). But the discussion of coreference is carried over to the final Chapter.

Then, in §4.3, I compare the theory developed here—called the 'edenic' theory of reference—to a recent and in some respects similar proposal due to Jeffrey King. The final section, §4.4, addresses an objection to the edenic theory, and, incidentally, to my critique of puzzle-driven semantics, based on recent work by Eliot Michaelson.

4.1 Edenic reference

As should be familiar by now, the Gricean program aims to explain the possibility of successful communication, and it proposes to do so by, in the first instance, *explicating* basic semantic and intentional notions. Linguistic meaning is explicated in terms of notions like *speaker meaning*, *intention*, and *belief*. *Saying* is explicated in terms of the coincidence or compatibility of speaker meaning and linguistic meaning on an occasion of utterance. Above I gave a rough Gricean explication of the foundational notion of speaker meaning. We have also seen, I hope compelling, arguments to the effect that *saying* and *referring* should be explicated in terms of speakers having certain specific communicative intentions and beliefs. Whether this constitutes a ‘revision,’ ‘addition,’ or even an ‘analysis’ of corresponding folk psychological notions is, in my view, beside the point. Explications take folk psychology or scientific theories—or combinations thereof—as starting point and are thus constrained to some extent, such that the explicated concept must bear some resemblance to the original concept that got us onto the problem in the first place.

In his 1982 paper, “Meaning revisited,” Grice writes:

[T]o say what a word means in a language is to say what it is in general *optimal* for speakers of that language to do with that word, or what use they are to make of it; what particular intentions on particular occasions it is *proper* for them to have, or optimal for them to have. Of course, there is no suggestion that they *always* have to have those intentions: it would merely be optimal, *ceterus paribus*, for them to have them. As regards what is optimal in any particular kind of case, there would have to be a cash value, an account of why this is optimal. (p. 299, first two emphases mine)

He then says that there can be a whole range of different accounts of this ‘cash value,’ mentioning that the usage might be conventional or it might be laid down by the inventor of an artificial language, but “what we get in every case, as a unification of all these accounts, is the optimality or propriety of a certain form of behavior.”¹

¹See, also, Grandy & Warner (1986: 25–26).

This is the spirit in which Grice's own explication of speaker meaning is to be taken and it (the explication) provides, in my view, the only non-circular account of linguistic meaning there is.² Along these same lines, I will now propose and argue for new ways to explicate *speaker reference* and *speaker coreference* within the Gricean framework. These are, to be sure, based on prior work by, in particular, Bach, Neale, and Schiffer.³ I see my contribution as additions to be made in reaction to the arguments developed in Chapters 1 and 2.

In §1.3 above, I introduced the Gricean (and Millikanian) idea of the proper or optimal function of singular terms as providing evidence for referential intentions. It is time to develop this idea in more detail. Following Neale's (forthcoming) discussion of Schiffer (1981), let's start by defining two related notions of reference, *referring in* and *referring with*.

Referring in (RI)

In (the course of) uttering σ , S refers to o if, and only if, in uttering σ , S means an o -dependent proposition.

Referring with (RW)

In uttering σ , S refers to o with e if, and only if, (1) e is properly contained in σ , and (2) $(\exists H)(\exists R)$ s.t. in uttering σ , S intends H to recognize that S was referring to o in uttering σ , at least partly on the basis of their mutual knowledge that $R(e, o)$.

The notion of *referring in* is the basic, ground-level notion of reference in Gricean theory. First, one can refer to o *in the course of* uttering σ without referring to o *with* any particular expression. This gives rise to *implicit* or *unarticulated* reference, as in Schiffer's example, cited above (p. 79), of uttering 'No' when asked 'Is Harry Truman alive?'

Secondly, notice that RI is derived directly from the prior notion of speaker meaning. On this view, the act of referring consist, in the first instance, in the act of speaker meaning a singular proposition. The notion of speaker meaning was explained in the last Chapter, but here I try to flesh out the notion of speaker meaning singular propositions. There is another tradition, broadly Strawsonian in spirit, according to which speaker meaning a (monadic) singular

²On this point, see Unnsteinsson (2014: 3344).

³Especially Bach (1987: 51–53); Neale (forthcoming); Schiffer (1981: 69–76).

proposition consists in two even more basic meaning-intentions. One is the intention to refer to a particular object, the other the intention to express or refer to a property which is then, by way of combining the two basic acts, predicated of the object referred to.⁴ My assumption here is that, on the contrary, speaker meaning a singular proposition is explanatorily prior to these two types of intentions or communicative acts.

Still, in what follows, I will be concerned with the more sophisticated act type of *referring to an object with a linguistic expression* and propose my own theory of such acts, starting with something akin to RW above. But RW has its problems. The notion of referring occurs on both sides of the biconditional and so the definition is not as illuminating as it could be. Also, the final part of RW is too strong in suggesting that *H*'s recognition needs to be based partly on the *mutual knowledge* that $R(e, o)$. Of course, R can represent any number of binary relations useful for communication (convention, salience, evidence, etc.), but it is not clear whether RW allows for cases where the speaker intends one such relation while the hearer's recognition is actually partly based on some distinct but equally useful relation. This would, it seems, be a case of successful reference and, so, the assumption of mutual knowledge of R appears to be too strong. It is better, instead, to explicate the notion of *referring with* in terms of the general notion of 'evidence.'

Gricean speaker w-reference (GSW)

Speaker S w-refers to o with e in uttering U iff U is an utterance of sentence type σ , e is properly contained in σ , $(\exists H)(\exists p)$ such that p is o -dependent and S utters U intending

- (1) to produce thereby in H the belief *that* p ;⁵
- (2) H to use the e -part of U as evidence that the belief in (1) concerns o ;
- (3) H to recognize S 's intentions (1) and (2).

To illustrate, suppose I utter 'Gill snores' intending thereby to produce in you the belief *that Gill snores*. There is, then, a proposition p with a *Gill*-dependent truth condition such that I

⁴For example, Strawson (1950: 17–19). See Bertolet (1987: 202–206); Schwarz (1976: 67–68); Stine (1978: 51–53) and, more recently, Hanks (2011); Harris (2014); Soames (2010b, 2014: 105). Soames speaks of the act types of *targeting* and *predicating*, combinations of which give rise to full propositional act types.

⁵Or the intention to make it such *that* p . But I ignore this qualification in what follows.

intend for you to believe p . My whole utterance is an event—involving me emitting a sequence of sounds or inscribing a sequence of letters—encompassing the subutterance associated with the subsequence /gill/ or ‘Gill.’ At its core, then, my *w-referential intention* (the second clause), on this occasion, consists in my intention that you use the ‘Gill’-part of my whole utterance as evidence that the belief in question concerns *Gill*. Note that there is always a co-occurring informative intention (1), so *w-reference* is never an isolated event (cf. Bach 1987: 51). And, further, when I perform an act of GSW I have both the usual informative-communicative intention with respect to (1) and the added referential-communicative intention with respect to (2), and this is made explicit in the third and final clause.

Admittedly, a full defence of GSW would appear to require a whole lot of argument and preemption of counterexamples. GSW is, as indicated briefly above, intended as a stipulative definition of utility in the construction of a semantic theory and not as a piece of conceptual analysis (cf. Schiffer *ibid.*, 68). And I believe more construction is in order before counterexamples are even appropriate.

The major thesis of this chapter is as follows. To successfully perform the speech act of GSW, the speaker must in addition satisfy a certain doxastic constraint:

Edenic constraint

Speaker S performs GSW at time t with respect to o only if there is no x such that, at t , S confuses x and o .

As expected, ‘confusing x and o ’ is defined as either *believing falsely that $o = x$* , or *believing falsely that $o \neq x$* , for any x . The mental state of confusion and the performance of GSW must be *simultaneous* in that the relevant false identity beliefs must be true of the speaker at the time of the speech act. The speaker must, explicitly or implicitly, hold the relevant beliefs at the time of utterance.⁶

When a speaker successfully performs the speech act of GSW, satisfying the constraint, we say that they refer *edenically* to the object in question. Such edenic reference presupposes a complex doxastic-intentional mental state which, in words reminiscent of Grice’s own, is

⁶See the Frege model of confusion, §1.2. Also the discussion, in §2.2.2, p. 54, about how lack of contextual salience can defeat the corrupting effect confusion has on acts of referring.

the *optimal state* with respect to referring, or if you like, to referring to an object with some linguistic expression (1982: 302). In many cases, of course, ordinary speakers may not realize exactly this complex state, and so they won't w-refer to *o*, in the strictest sense, on those occasions. GSW plus the edenic constraint still captures the mental state proper to w-reference. Or so I argue.

Surely there are many doxastic states involved in any Gricean explication of speaker meaning and speaker reference, but it's customary to formulate them as constraints on communicative intentions themselves. For example, Griceans endorse the Humpty Dumpty constraint which says, roughly, that speakers cannot intend what they believe to be impossible.⁷ The edenic constraint has a slightly different function. Speakers can form intentions to refer or think about an intentional object *o* even if they have false identity beliefs about *o*. If Millikan is right, this manifests a special kind of cognitive corruption, but the very formation of these corrupt thoughts or intentions is not precluded. Only the performance of a particular communicative act is precluded.

This may seem to go against one of Grice's fundamental insights. In *Relevance*, Sperber & Wilson write that his "greatest originality" was not to suggest that communication involved intention-recognition. That's mere common sense. It was, they argue, "to suggest that this characterisation is sufficient: as long as there is some way of recognising the communicator's intentions, then communication is possible" (1986/1995: 25). I think this is substantially correct and important, although it misleadingly places all emphasis on the hearer's side of the equation. It has often been argued that Gricean models of communication are too simplistic.⁸ I have a thought, pick a string of words which will most likely, by my own lights, produce this very thought in your mind; you hear the words and, by some way or other, the thought is replicated in your head. But the model is supposed to be a simplifying idealization, and that's where much

⁷See note 25 in Chapter 1, p. 32.

⁸See, e.g., Buchanan (2010). A group of theorists unsympathetic to the Gricean program label this the 'Lockean' or 'communicative' conception of language, arguing that it is naïvely mentalistic and in need of significant revision or, more strongly, replacement. See, in particular, Bar-On (2013); Carruthers (1996); Gauker (2003); Green (2007). These authors are united in the belief that utterances express mental states directly, rather than expressing intentions to produce an effect in an audience. See also Rosenthal (1986, 1989) for an early comparison of such a theory to the Gricean view.

of its explanatory power comes from.⁹ The model characterizes communication when all goes well; when the mental states of speaker and hearer are optimal with respect to communicating contents. And when the speaker's perspective is taken fully into account, it will be seen that intention-recognition by itself is not sufficient to explain communication.

To be clear, then, the notion of edenic reference is intended as an explication of a specific type of speech act in terms of the doxastic-intentional state of the speaker. The hearer enters this picture only in terms of the beliefs that the speaker has about the epistemic situation of the hearer: beliefs that place constraint on the kinds of communicative intentions that the speaker can actually form. As Grice writes it in the quote above, we are concerned with the optimality or propriety of a "certain form of behavior." I argue that edenic reference is explanatorily and metaphysically basic in giving a theory of meaning and communication in natural language. What follows is a list of four arguments. The first two briefly recapitulate arguments from Chapters 1 and 2, clarifying their relevance to the present discussion; the others provide additional support.

4.1.1 Confusion corrupts the evidence

In Chapter 1, I argued that the mental state of confusion disrupts the proper function of singular terms. As already suggested, there is significant overlap between Gricean optimal states and Millikanian proper functions.¹⁰ In a nutshell, the point was that whenever a speaker's utterance contains a subutterance U_{sub} of a *confused* singular term e , the speaker's referential intention is *conflicting* and the evidence provided by U_{sub} is, thereby, corrupt. The intention is 'conflicting' in that either (i) e is really anchored equally in two objects while uttering e is 'supposed' to indicate only one, or (ii) e and another expression e' are really anchored in a single object while uttering e is 'supposed' to indicate only something distinct from what would be indicated by e' . The evidence provided by U_{sub} is 'corrupt' in (i), i.e. combinatory confusions, for it optimally functions to pick out a single object while picking out two. In (ii), i.e. separatory confusions, the evidence functions optimally to pick out a single object while picking out none, for, in the relevant cases, nothing is picked out by e while not being picked out by e' as well.

⁹On the use of models in science and philosophy of mind, see Godfrey-Smith (2005, 2004).

¹⁰See, again, Origgi & Sperber (2000) for discussion.

That's an awfully condensed way of putting the whole matter, so let me clarify with a brief, but admittedly limited, analogy. Consider Tyler Burge's (1979) arthritis case. The protagonist *S* believes many things about *arthritis*: *that arthritis is uncomfortable, that she has arthritis in her wrists*, etc. But *S* also has a few false beliefs. *S* believes

(45) *that she has arthritis in her thigh*,

and, accordingly, *that arthritis can be something other than specifically an inflammation of joints*. In Burge's original story, *S* reports to a doctor her belief (45) and is corrected. But imagine, instead, that *S* simply utters

(46) Arthritis is annoying.

For sake of argument, assume, first, that *S*'s reference to *arthritis* is not somehow 'deferential'—to community, or to experts¹¹—and, second, that in (46) *S* refers to her own personal *arthritis*-concept, which is just like the concept of everyone else in *S*'s linguistic community, except that its application is not restricted specifically to *joints*. *S*'s concept properly applies to certain ailments in the thigh, for instance.

Call *S*'s deviant concept *arthritis** and suppose, then, that by uttering (46) she speaker-means (47).

(47) *that arthritis* is annoying*.

I hope the moral to be drawn is obvious by now. I would claim, holding the various assumptions in place, that if (47) indeed specifies what is said by *S* in uttering (46) on a given occasion, then (46) is misleading as a piece of linguistic evidence. Granted, if false beliefs always engender this sort of corruption it is awfully common. Furthermore, it seems like the practical upshot would often be negligible. The fact remains that in *S*'s linguistic community (46) is an ideal piece of evidence for the content in (48) and *not* for (47).

(48) *that arthritis is annoying*.

¹¹For a healthy dose of skepticism about the notion of deferred reference, see Greenberg (2007, 2014).

It is easy, however, to concoct cases where conceptual mismatch of this sort would be of considerable consequence. A doctor who hears *S*'s talk of her annoying arthritis would probably be misled into looking in the wrong places. Surely this can create serious problems in individual cases.

It is not my intention to argue that cases like this should be idealized away in exactly the same way as the ones I've focused on, and I'm not sure this should be done. That's a story for another occasion. I recognize, however, that it's possible to describe even the arthritis case as involving a false identity belief about *arthritis* and *arthritis**. But, for one thing, the idea of deferential reference is more plausible for the arthritis case than for the others. Such a move would block any more substantive analogy between the two kinds of case. Anyhow, the purpose of this exercise has just been to establish a similar case where the corruption of the linguistic evidence seems obvious.

4.1.2 Puzzles are problematic

This was one major lesson from Chapter 2. For example, I found fault with semantic reference on the grounds that it was unnecessary in giving a plausible explanation of Kripke's assumption. But more generally, any semantic entity *X* is theoretically illegitimate if *X* is postulated only on the grounds that it helps solve puzzles involving confused agents. Surely one can imagine other and better reasons for postulating something like semantic reference, or Fregean senses for that matter. If, however, *X* is introduced only in order to explain how *confused* agents can successfully communicate with each other the evidence for *X* cannot be sufficient. One must ask, for example, whether there is any reason to suppose that the explanation of successful communication between edenic speakers also requires an appeal to the category of *X*'s.

This suggests, again, that GSW and the edenic constraint jointly provide a reasonable explication of speaker reference.

4.1.3 Edenic reference is part of pragmatic competence

Competent speakers implicitly believe, in virtue of their linguistic competence, that confusion ought to be avoided if one is to communicate successfully. The edenic constraint thus captures an interesting fact about pragmatic or communicative competence. An adequate theory of the speaker-hearer's linguistic competence will include at least three sub-theories; (i) a theory of syntactic competence, (ii) a theory of semantic competence, and (iii) a theory of pragmatic competence. This last part describes the mechanisms in virtue of which speakers are able to plan their utterances in light of their beliefs about the mental states of the audience and, correlatively, how hearers are able to correctly attribute communicative intentions to speakers in specific contexts. Pragmatic competence is thus a rich system of inferential mechanisms and capacities for mindreading. As Chomsky (1980: 225) puts it, "pragmatic competence places language in the institutional setting of its use, relating intentions and purposes to the linguistic means at hand."¹²

It is important to remember, from Chapter 1, that I am explicitly abstracting away from global confusions, such as when scientists are entirely incapable, over a long stretch of time, of distinguishing two natural properties by using any of the means available to them. Local confusions, however, are always such that the confused person is epistemically blameworthy, even if only to a very small degree. For example, even if the deception is quite elaborate and well executed, Lois ought to be able to disabuse herself of her confusion about Superman using mere pedestrian powers of discrimination.

Patterns of ordinary criticism provide evidence for edenic reference being part of pragmatic competence. Examples like the ones invoked by Soames in support of his theory of anaphora are perhaps the most striking in this context. To repeat one of them, Lois says to *H*,¹³

¹²See also Carston (1998: 5–7, 2002: 10–11); Neale (2005: 188–189). Sperber & Wilson (2002a) argue that what I'm calling pragmatic competence calls for a specific mindreading module in the mind geared towards maximizing the overall relevance of utterances. Earlier relevance theory had been presented as describing cognitive systems belonging to domain-inspecific central processes. I mention this only to point out that, so far as my claim about conflicting referential intentions goes, I don't need to take a stand on this issue. Furthermore, the rules or mechanisms that make up pragmatic competence can either be brute and "merely embodied" or fully representational. See Devitt (2006: §3.1) on that point.

¹³See page 66 in Chapter 2.4.

(24) Superman loves his_↓ mother.

And she points to Clark Kent—i.e. Superman wearing disguise—intending ‘his_↓’ deictically, as signalled here by the downward arrow. Add that everyone except Lois now knows the truth and there is no reason at all for *H* to keep her out of the loop. In such a case *H* is justified in correcting Lois: ‘You’re confused. His_↓ mother is Superman’s mother. He_↓ is Superman.’

After accepting the correction Lois might say ‘Right, so what I really wanted to say was that (49).’

(49) Superman¹ loves his₁ mother.¹⁴

Where the number 1 signals some suitable relation of binding between expressions.

Remember also the example on page 22 in Chapter 1, of someone who combinatorily confuses Bill and Biff, calling both ‘Bill.’ When such a speaker makes the mistake of greeting Biff with ‘Bill’ a hearer may offer a polite correction because, in the jargon of the theory, this act of *w*-reference wasn’t edenic. Of course the folk would put it differently. If this is right confused reference is, as already argued, a special kind of performance error—similar to malapropisms.

Thus, the speech act of referring is subject to, as we might call it, the rule or maxim of edenic reference. Mastering the act of referring involves, among other things, becoming sensitive to the maxim. Of course, competent speakers do not consciously represent the maxim, but they behave as if in accordance with it, particularly when judging some acts of reference to be proper and others improper or somehow faulty. The maxim can be formulated like this.

Edenic maxim

Try to refer to *o* with *e* only if you thereby refer edenically to *o*.¹⁵

¹⁴Remember that in some languages, such as Icelandic, (24) and (49) have different translations.

¹⁵Here I am indebted by Timothy Williamson’s (2000: 243) so-called knowledge norm of assertion. I do not, however, endorse his view that the knowledge norm is constitutive of assertion—but it is part of speakers’ semantic or pragmatic competence. Similarly, if the reference rule is constitutive of anything it is merely constitutive of the theoretical notion of *w*-reference, which is also part of pragmatic competence.

To be clear, the claim here is only that speakers who flout the maxim are, by definition, unable to perform the act of ‘edenic’ reference, but they can still manage to perform an act of reference, in an ordinary sense of the word. It is just that the act is deficient in that it breaks this implicitly accepted rule. By uttering a singular term—especially if it is a proper name—one normally, but only implicitly, *represents oneself* to one’s interlocutor as not having any false identity beliefs about the object referred to. One represents oneself as uttering a token that is causally grounded in the referent.¹⁶ And of course one can be wrong in so representing oneself.

4.1.4 Edenic reference is explanatorily basic

For now, let’s focus only on proper names as devices of reference. What I’ll say here can be extended to other referring expressions, but this would involve us in unnecessary complications. I hold the following thesis about how to explain the successful introduction and perpetuation of the practice of referring with a proper name in a population of speakers.

Happy names

For any name n , object o , and group of speakers G , if the act-type of uttering n is to stabilize as providing good evidence for the intention to refer to o in G , the total number of false identity beliefs about o in G must stay below some critical threshold.

Names are happy if they satisfy the condition, unhappy if they don’t. The condition is satisfied if a sufficient number of n -utterances are edenic references to o . What is a sufficient number? The threshold will certainly vary relative to contexts and the interests of speakers. The extreme cases are most interesting, so all we need for now is the idea that at some point or other confusion about the identity of o will make names of o unhappy. By definition, an unhappy name does not perform its proper function of evidencing a referential intention.

All this may seem too obvious to bear spelling out. But let me explain why the notion of an unhappy proper name ought to be surprising to theorists of a broadly externalist persuasion such as myself. One major insight of externalism in the philosophy of language and mind was that reference could be secured even in the face of quite extreme types of speaker ignorance

¹⁶Cf. Evans (1982: 310); Williamson (2000: 252n6).

and error (e.g. Donnellan 1970; Kripke 1980; Putnam 1973; Wittgenstein 1953: §79). And the insight constitutes a compelling objection to some simple varieties of descriptivism about names.¹⁷ To illustrate, even if some of the ancients seem to have believed falsely that the Earth was flat and at the center of the universe we have no problem attributing beliefs about the Earth to them. They presumably believed *that the Earth was below their feet*, where the object of this belief is identical to the object of latter-day Earth-beliefs (Devitt 1984/1991: 159–160).

Thus there is a distinction between extreme cases of confused identity and other cases of substantial ignorance and error. A group of three speakers where the majority of each speaker's Earth-beliefs are false and no two of them share the same false Earth-beliefs would still be able to communicate with each other about the Earth. If we picture the same situation except that the beliefs in question are all false *identity* beliefs the practice of using the name wouldn't get off the ground—the name wouldn't function at all in interpersonal communication.¹⁸ Therefore, I suggest, even on broadly externalist assumptions there is a type of error, namely confused identity, that makes reference impossible in certain cases. As in §4.1.1 above, however, we must assume that the confused speaker can't be saved here by some notion of deferred reference.¹⁹

Going back to the Millikanian notion of normal explanation outlined in Chapter 1, we can say that the happiness of name *n* of *o* in group *G* is explained by the historical series of edenic *n*-utterances in *G* such that uttering *n* on those occasions provided good evidence for an intention to refer to *o*. Past *n*-utterances made by confused members of *G* are excluded from the series and thus don't contribute to explaining the stabilization of the *n*-using practice. This is the sense in which the edenic constraint pinpoints an explanatorily basic category of referring with a name; the constraint must be satisfied if the name is to perform its proper evidential function and its continued use is to be explained. Edenic reference is the 'normal' condition of names, in Millikan's sense of the term.

¹⁷But apparently not to others, see, e.g., Bach (1987: 157–159).

¹⁸I see Devitt (1974: 201, 1981a) as making a very similar point.

¹⁹See, again, Greenberg (2007, 2014).

4.2 Coreferential cognizance

In addition to edenic reference there is a related constraint on acts of uttering sentences containing more than one singular term. The basic idea here is that in order to properly perform such acts the speaker must—if only tacitly—make some assumption or other about the coreference or noncoreference of distinct singular terms, unless one of the terms is actually *bound* by the other. Indeed, the standard view in syntax is that bound, coindexed expressions are covalued or coreferring in virtue of grammar or linguistic meaning (Fiengo & May 1994). Noncoreference is ruled out by grammar.

Cognizance constraint

If S utters sentence σ intending to express the singular proposition *that* p , and σ contains singular terms e_1, \dots, e_n , then S must not be *indifferent* as to whether S intends, by uttering any two expressions e_i and e_j in σ , to corefer to a single object or to refer to distinct objects.

Note that the constraint does not apply across different whole utterances or distinct propositions. The hypothesis is, then, that intrasentential or intraclausal cognizance is part of pragmatic competence but extrasentential cognizance is not. To illustrate, if a speaker never has any opinion about whether what she referred to on one occasion is identical to what she referred to on a separate occasion she can still have full pragmatic competence. To take an extreme example, she might be a perdurantist and believe that objects are constituted by non-identical temporal parts, insisting that she always only refers to such parts, never the objects themselves. According to the cognizance constraint, however, even the obstinate perdurantist will form intentions to refer to the same or to different such parts in expressing a single proposition. Occurrences of pronouns bound by a name in subject position is a particularly clear example and the cognizance constraint may help explain the intuition that such occurrences aren't genuinely 'referential': but of course they are in the sense that they function—together with the name—as evidence for a referential intention.²⁰

²⁰Thus GSW can be restated for the notion of *coreferring with an expression*: Speaker S corefers to o with e in uttering U iff U is an utterance of sentence type σ , e_1, \dots, e_n are properly contained in σ , $(\exists H)(\exists p)$ such that p is o -dependent and S utters U intending (1) to produce

However, indicated above, the constraint only applies to singular terms that are not bound by a c-commanding singular term or quantified determiner phrase. In those cases the speaker can indeed express, when there is only one binding expression and one bound, a monadic proposition that can be represented via λ -abstraction. So, a speaker who utters (20)—such that the occurrence of ‘his’ is genuinely bound by the occurrence of ‘John’—or (50) expresses (51) or (52) respectively (see, e.g., Neale 2005: 205). Note that I am only considering the ‘sloppy’ reading of these sentences here, ‘strict’ readings are discussed in the next Chapter.

(20) John loves his mother

(50) Every man loves his mother

(51) $\lambda x \langle x \text{ loves } x\text{'s mother} \rangle \text{John}$

(52) [every x : man x](x loves x 's mother)

Thus, it should be clear, bound, sloppy-reading occurrences are not really referential occurrences and, so, they do not involve any kind of coreference to which the constraint applies. In Chapter 5 below, however, I shall argue that edenic and cognizant speakers who utter coreferential sentences not involving binding or explicitly reflexive constructions can still express, thereby, monadic propositions like the one described in (51).

If we focus only on sentences with two purportedly unbound singular terms, the constraint says that performing an act of uttering such a sentence requires the speaker either to intend the terms to corefer or not. In some cases, however, mentioned here only so they can be left to one side, the speaker's intention will have to take scope over the disjunction. For instance, I may utter ‘I wonder whether Cicero is Tully’ and, then, (part of) my intention would be *that either the names corefer or they do not*, but I don't know which. Otherwise the performance is infelicitous. In what follows I present one argument for the cognizance constraint. The next Chapter continues this line of argument, providing additional support.

thereby in H the belief *that* p , (2) H to use the e_1, \dots, e_n -parts of U as evidence that the belief in (1) concerns o , (3) H to recognize S 's intentions (1) and (2).

4.2.1 Do you know what you're saying?

When speakers violate the cognizance constraint it is reasonable to conclude that they don't actually know what they (will be taken to) say and mean. Part of my argument against Soames in §2.4 above made use of exactly this point. Mary is coreferentially indifferent if she is, say, blindfolded when she utters

(20) John loves his mother,

and points at random to indicate the referent of the third person possessive pronoun. Sure, there is a highly general description of her intention even here. It's true, for example, that Mary meant *that John loves the mother of whatever object or individual she appears²¹ to point to while blindfolded*. Normally, however, knowing what one says in uttering something like (20) requires a more specific identification of the referent (cf. Evans 1982: 171, 316; Kaplan 1978: 390).

Admittedly, the ordinary notions of 'knowing what one said' or 'knowing who one referred to' are vague (cf. Hawthorne & Manley 2012: 71–73). But the cognizance constraint exploits a narrower notion, namely that of intended coreference or noncoreference. Let me explain. If a speaker *S* refers to *o* with a single occurrence of a singular term *e* in a simple monadic sentence *S* can be indifferent about which object she in fact (was taken to have) referred to, and still know what she said. To illustrate, imagine that *S* is with a friend and walks past a couple of strangers and hears one of them utter the name 'Ursula.' *A* can later say to her friend,

(53) He was talking about Ursula.

Even with this little information the speaker seems able to competently communicate a proposition about *Ursula*. If needed, the speaker can explain later what she intended, and so on. I am assuming that the talking stranger is salient enough in the context for *S* to intend to refer to him with 'he.'

Most likely, however, when *S* utters (53) she intends 'he' to refer an object distinct from the referent of 'Ursula' in (53). Reasonably enough, *S* will believe that the stranger is talking

²¹I say 'appears' because *pointing* is clearly an intentional notion: one doesn't really point to something just by aimlessly waving a finger in the air. See note 28 in Chapter 2, p. 55.

about someone other than themselves. There will be a shared presumption between *S* and her friend to the effect that *S* is coreferentially cognizant in uttering (53), i.e. that *S* either has a coreferential intention or a noncoreferential intention. If *S* would turn out to have been coreferentially indifferent—e.g. if she adds, ‘But maybe that was Ursula talking about herself’—there is a more precise sense in which she doesn’t know which proposition she intended to express by uttering (53). She neither intended to refer to two objects nor did she intend to refer to a single object

Quite generally, uttering an expression of the form *Rab* places the following constraint on the competent speaker’s formation of a communicative intention: The speaker must either believe *that a = b* or *that a ≠ b*. Otherwise the speaker makes a pragmatic performance error. The error even seems to give rise to a Moore-like paradox, similar to ‘It’s raining but I don’t believe it is.’ Such Moore-paradoxical sentences are often called “pragmatic contradictions” in the sense that what one strongly implicates or presupposes with the first conjunct one negates with the second.²² And even if implicatures are cancellable some cancellations are just plain absurd. So, consider (54).

(54) Ruth and Benazir are here, but I believe they are one and the same person.

Now, surely, Moore-paradoxical sentences vary in levels of absurdity and it seems quite possible to dream up contexts in which something like (54) is appropriate. Think of Lois in the midst of discovering the identity of Superman, say. But the same thing can be said about Moore’s original examples. Just imagine a neurological condition where perception and belief are somehow disconnected and the patient actually realizes this. My only claim, here, is that an utterance of (54) could easily be a kind of pragmatic contradiction, since uttering the first conjunct strongly suggests or presupposes, in typical contexts, that the speaker believes the negation of what is affirmed by uttering the second conjunct.

²²See, e.g., Shoemaker (1995: 222). According to relevance theory what is said by uttering the second conjunct may be a higher-level explicature of what is said by uttering the first (Sperber & Wilson 2005: 23–24).

4.3 The edenic theory and King's account

To give it a name, call the theory propounded here the edenic theory of reference. It can be summarized roughly as follows. When an edenic speaker *S* utters a singular term *e* on some occasion the referent of *e* is the object *S* intends to refer to with *e* on that occasion. If *S* is confused with respect to the term and object, the theory doesn't predict that there is a unique object successfully referred to, but there will invariably be more or less useful interpretations of the speaker's utterance from the hearer's point of view—such as taking the speaker to refer to the object that an edenic speaker *would have* referred to by making the same type of utterance in the same context. Depending on the facts about the confused speaker's mental state, the intended referent will either be two distinct objects or no object at all. For more complex sentences the speaker must also be coreferentially cognizant.

I have mostly been concerned with proper names as evidence for referential intentions, but demonstratives and indexicals fall under the purview of the edenic theory as well. Now suppose that *S* utters,

(55) That's a human being

referring to a small speck in *S*'s visual field that's so far away that it looks like a strange rock-formation. The speck is really a person. If *S* is Lois Lane and the speck is Clark Kent, then she's confused and the edenic theory predicts that her utterance cannot reach the standard required for speaker *w*-reference. And if she refers to someone *else* who's also a speck in her visual field at the same time, someone she has never referred to before, then her utterance of (55) does satisfy the standard. This consequence is radically unintuitive especially since the cases can be subjectively indistinguishable. But one difference between simple demonstratives and proper names is that utterances of the former carry a different commitment: the speaker doesn't suggest that the object is a bearer of a name uttered. This makes it easier for the interpreter to abstract away from the perspective of the speaker and assign an edenic act of reference to the speaker. Compare a variant of Kripke's case where only the speaker is confused and the hearer is actually aware of the confusion. In that case it's not even clear what the most reasonable interpretation should be. Which is better for the hearer; to take the speaker as

referring to Smith or as referring Jones? No principled answer can be given I think. But when the only indication is a simple demonstrative the case is very different from the standpoint of reasonable interpretation. If the guy over there is Smith, it is most reasonable to take the speaker to be referring to Smith with the demonstrative.

Furthermore, if we add a Reimerian distinction between primary and secondary referential intentions we can say that in the the demonstrative case the speaker's intention to refer to the person in the distance has more weight than the intention to refer to someone who's falsely believed to be identical to Smith and to Jones (see §2.2.2 above).

The edenic theory is similar, in some respects, to a view recently defended by Jeffrey King (2013, 2014). According to his 'coordination account' of demonstrative reference an object *o* is the semantic value of the use of a demonstrative on an occasion if two conditions are satisfied: (1) The speaker intends to refer to *o*. (2) A competent, attentive, reasonable hearer *would take* *o* to be the object intended. Thus the reference must be such that an ideal interpreter would produce the correct interpretation. The two conditions are necessary for reference to occur, so, if either one doesn't hold in a given case, the demonstrative gets no object assigned as a semantic value. Much of the motivation for the coordination account can be adduced in support of the edenic theory and vice versa. Both proposals argue for idealizing the speech act of w-referring.

There is, however, one very clear reason why the edenic theory should be preferred to the coordination account. King's second condition obscures the goal of communication since that goal cannot be defined, even partly, in terms of an ideally rational hearer. The goal of the hearer is to discover the speaker's communicative intention, full stop. It is not to discover how a competent and reasonable hearer *would* interpret the speaker on the occasion of utterance. Here's one way to see this. Suppose *B* is not competent, attentive or reasonable. Suppose also that *A* knows all of this about *B*. Then *A* utters σ intending to refer to object *o*. It should follow that, in such a case, this interpretation is *not* one which a competent, attentive, and reasonable hearer would come up with. The utterance is specifically designed by *A*, on grounds of *A*'s beliefs about the incompetence of *B*, for a hearer that's not so constituted. Of course, however, communication succeeds and *B* understands that by uttering σ , *A* said something about *o*. And,

further, o is objectively what A referred to on this occasion.

The problem is simply that the coordination account, as here understood, makes a different prediction. For presumably it is possible, on this view, that a competent, attentive and reasonable hearer would have interpreted A as referring to o' , where $o \neq o'$. The account then predicts that the demonstrative has no semantic value on the occasion of utterance. But this is absurd. Furthermore, if King's account is changed—for example by relativizing the definition of a competent speaker/hearer to a given context—to allow that, in a case like this, o is in fact what A referred to, it would tend to make the second condition superfluous. The second condition would, then, tend to state merely the hearer must interpret the speaker *correctly*, i.e. grasp the speaker's actual intent. But surely the theory must allow that a speaker actually refers to an object without the hearer getting it. I conclude that King's coordination account is a non-starter. And, finally, the edenic theory has no analogous problems, since the idealization is only on the speaker's side of the equation, skipping the ideally rational hearer completely.

4.4 Objection: Intentional confusions

Eliot Michaelson (pers. comm., see also his 2013: ch. 4) provides an interesting objection to the edenic theory and, equally, to my doubts about puzzle-driven arguments in §2.2 above. Here is how I understand the point. We can imagine a case like Donnellan's and Kripke's where the speaker has not really confused the identity of the referent but yet the same implications would seem to follow, i.e. there will be the same kind of conflict in referential intention and the same type of corruption of the linguistic evidence. This will be a case of intentionally confused or deliberately obfuscating utterances. Therefore the Frege model does not really explain intentional conflict in the way required by my theory. And, further, the edenic constraint on reference would have to incorporate a whole variety of new cases; perhaps so much so that it becomes unwieldy and overly general.

You and I are taking our usual early morning stroll. We see a man in the distance. I know it is Smith, our mutual friend. I know also that you have not recognized the man yet. I decide to deceive you into believing that the man in the distance is not Smith but Jones. Jones is another acquaintance of ours. So I utter

(10) Jones is raking the leaves.

Assuming that I succeed, you come to believe

(56) *that Jones is raking the leaves*

By now you are truly confused, since you believe that Jones is the man over there and the man over there is in fact Smith. Thus if you use 'Jones' in a sentence in this context your referential intention will be conflicting.

But what about my own original use of the name in (10)? I do not believe that Jones is Smith. I know full well that I am referring to Smith in uttering (10). But I am, surely, also referring to Jones because I intend to produce in you the belief (56). Thus we have a case where the speaker's referential intention is conflicting without the speaker having false beliefs about identity.

The example should remind one of the discussion of intentional malapropisms in Chapter 3. In a similar manner, I don't think this kind of case provides occasion for significant disagreement between theorists. A speaker who utters (10) in this context is doing at least two things at the same time. First, he intentionally but not explicitly introduces a new name for an individual, i.e. Smith. In Perry's (2012: §6.3) terms he instigates a permissive convention: the interlocutors are now allowed to use 'Jones' as a name for the man raking leaves. This succeeds only if the convention is endorsed—implicitly or explicitly—by the hearer. If the hearer's confusion persists, she will not have realized that a new name-using practice has in effect been introduced. Since the speaker knowingly and deliberately uses the name in a manner that has not been established in any relevant idiolect it is undeniable that she is introducing a *new* convention. It is not customary, for the speaker and hearer, to utter 'Jones' to intentionally refer to Smith.

This sort of thing happens all the time. Just picture yourself listening to someone who is an expert in a field you know nothing about. The expert will knowingly make utterances by which she both introduces a new name into your idiolect and says something about the object thus named. A botanist may point to a plant and say, "*Pinguicula vulgaris* here is a carnivorous plant." In this case, of course, the practice was already part of the speaker's idiolect. Importantly,

however, introducing new expressions can be more or less explicit, more or less misleading, obfuscating, and so on.

This brings us to the other thing the speaker is doing by uttering (10) in this particular context. He deliberately induces the hearer to believe that ‘Jones’ is being used in a way that is entirely familiar rather than novel. A convenient way of describing this is to posit two homophonic names ‘Jones₁’ and ‘Jones₂.’²³ ‘Jones₁’ is the name just introduced for the man over there, i.e. Smith, and ‘Jones₂’ is the name of our friend Jones in our shared language. In these terms, the second aim of the speaker is to utter ‘Jones₁ ...’ while producing in the hearer the belief that ‘Jones₂ ...’ was actually uttered. And surely this makes the example quite atypical on its own. It was not my contention that abnormal reference is only engendered by properly confused speakers. My argument was that such cases were paradigms for the puzzle-driven tradition in semantics.

One advantage of the foregoing analysis is that it explains a range of appropriate responses to (10). For example, if the hearer is not fooled and realizes exactly what the speaker was trying to do, she might say many different things, but on my account, these are both felicitous:

(57) Yeah, sure, Jones₁ over there is raking the leaves.

(58) That’s Smith. Jones₂ is in Luang Prabang.

In (57) the hearer is implicitly endorsing the new convention, at least temporarily, and with a touch of irony—as in: *you didn’t fool me*. The speaker might even say right away, with a suggestive intonation on the name: ‘Jones₁ has really changed since I last saw him.’ By uttering (58) the hearer rejects the new name.

Consider the expert botanist again. Call him Boaz. He spots a beautiful *Lamprocapnos spectabilis* in the distance and says to you,

(59) I see a bleeding heart on the ground there.

Clearly, Boaz can utter (59) intending ‘bleeding heart’ as the name of a plant while, in the same

²³Talking in terms of two distinct names is only a simplification here. It would probably be more accurate to talk of one name with two bearers. The speaker would still be introducing a new convention, namely the convention that Jones’ name can now also be used to refer to Smith.

breath, intending for you to believe—if only for a second—that there is an actual animal heart there. This is intuitively described in terms of two different linguistic expressions.

It should be plain by now that there is one crucial difference between genuine confusion cases and deliberate obfuscation of this kind. In the latter, one can always distinguish the obfuscating aspect and the non-obfuscating aspect of the referential intention. In (59) Boaz's obfuscation consists in intending to produce in you the false belief that there is a bloody heart there. The non-obfuscating aspect is that he intends to introduce the name 'bleeding heart' for *Lamprocapnos spectabilis* into your idiolect.

The original case with Smith and Jones is different only in that the newly introduced name-practice is new to the idiolects of *both* speaker and hearer. At least in the paradigm cases of speakers with false identity beliefs where this results in conflicting intentions, neither aspect of the intention is somehow more accurate than the other. And, again as in the case of intentional malaprops, the Gricean should be happy to explain these cases with a distinction between saying and meaning because, here, both are determined by the speaker's communicative intention. So, perhaps, Boaz intentionally said *that there was a bloody heart there* while also meaning *that there was a flower called bleeding heart there*. Admittedly, this description applies to the Smith-Jones case only if the speaker ultimately wants the hearer to take the primary meaning intended by the utterance to be *that Jones₁ is raking the leaves*.

In closing, I agree that deliberate obfuscations are interesting cases but also think they are abnormal in much the same way as cases of confused identity. But they are also importantly different, and definitely have not been used in the puzzle-driven tradition to draw illicit semantic conclusions. In particular, they do not provide reasons to think that what is said is not determined by speaker meaning (and its being compatible with the linguistic meaning of the sentence uttered).

Conclusion

The explication of speaker reference proposed here is designed to provide support for an insight that Strawson expressed and Neale (2005: 181–183) and others have kept alive. Here is Strawson, making the point in terms of descriptions rather than singular terms:

[W]here (i) there is an item which the speaker intends to refer to and that item answers to the description he uses, and (ii) there is also an item which answers to that description and of which it is true that in the social and physical context of the speaker's utterance it would be reasonable and natural to take it that a speaker, speaking conventionally in that context, would mean that item, and (iii) these items are not identical, then it is the first item and not the second which the proposition asserted by the speaker is about: so that in such a case, it is to be what the speaker actually *meant*, rather than what he would normally be taken to mean, that governs the ruling on what he *said*. (1974: 52)

Puzzle-driven semantics flourishes when theorists draw a conclusion opposite to Strawson's. The reason being that whereas confusion makes it difficult, at least in extreme cases, to fix exactly which content was intended by the speaker on a given occasion of utterance, it is always possible to identify some interpretation that would be 'reasonable and natural,' in the context, to attribute to the speaker. But these are importantly different phenomena. One is constitutive of the content of what the speaker says on a given occasion of utterance, the other is an epistemological ideal that hearers try to reach when engaged in interpretation—although their primary objective is simply to discover the speaker's actual intent.

The edenic theory follows Grice's advice in "Meaning revisited," where he writes that a full account of speaker meaning calls for the *absence* of certain 'sneaky' intentions (1982: 302–303). According to the edenic theory false identity beliefs and coreferential indifference give rise to sneaky—although not deliberately so—referential intentions, the absence of which needs to be stated in a full theory.

Chapter 5

Representation without thought

Introduction

In this Chapter, I develop and argue for a new way of thinking about the logical form of coreferential utterances. First, in §5.1, I explain the blueprint theory of linguistic meaning, according to which sentences are associated with highly general instructions for message construction. This idea sits well with the edenic constraint from Chapter 4, in that the constraint is, arguably, built into blueprints associated with sentences containing singular terms.

But what exactly is the connection between the cognizance constraint and blueprints associated with sentences containing more than one singular term? To answer this question I formulate a thesis called the ‘monadic thesis,’ which states, roughly, that cognizant utterances of coreferential constructions normally or optimally express monadic contents, even if the surface form of the sentence is polyadic.

The bulk of the Chapter is, then, devoted to the presentation of four arguments in favor of the monadic thesis (§5.2). They are, in different ways, motivated by the idea that coreference is always partly explained by the arbitrariness of linguistic forms, in the sense that one can invariably posit an equivalent linguistic expression where one of the singular terms is eliminated. Stated in this way, the claim is more or less trivial. But the arguments are designed to show that the proposition primarily expressed by the speaker can be monadic even if the sentence uttered happens to be polyadic.

Finally, in §5.3, I show that, if the monadic thesis is true, it gives rise to a new kind of structural mismatch between linguistic meaning and propositional content, mismatch which cannot be captured in terms familiar in the literature on context-sensitivity and underspeci-

fication. I argue, further, that intention-based semantics, and the attendant blueprint theory of linguistic meaning, have the potential to better explain this kind of mismatch than other theories, for example minimalism and hidden indexicalism.

To explain the import of the monadic thesis in more detail, I also compare the cases under discussion to other types of repetition or redundancy in natural language. The point being that the coreferential cases philosophers and linguists have been most interested in—identity statements, indexicals, reflexives, etc.—are linguistically highly entrenched phenomena of a more general kind. The class in question, which I call ‘structural repetitions,’ also contains appositives and complex singular terms.

5.1 Blueprints and coreference

According to intentionalism, as defined here, linguistic meaning places constraints on communicative intentions. Griceans differ, however, when asked about the exact nature of these constraints. I will assume, with Neale (2005: 189–192) and Schiffer (2003: ch. 4, 2014), that the linguistically encoded meaning of an expression is a non-propositional *blueprint* or *template* on the basis of which full-blooded propositions can be constructed.¹ This view harks back to Strawson’s old picture of linguistic meaning as encoding *general directions* for use.² In earlier Chapters I have described meaning in more general terms as part of the overall evidence provided by the speaker on the occasion of utterance, for their communicative intention.

Every meaningful sentence type in a given language is associated with a blueprint specifying a type of speech act and a type of proposition. Going along with Schiffer, we can represent the blueprint as an ordered pair,

$$(60) \quad \langle \vdash, \Psi \rangle$$

Where ‘ \vdash ’ stands for a type of speech-act or illocutionary force and ‘ Ψ ’ for a proposition type.

For example, if the sentence type is

¹Bach’s (1994a: 127) propositional radicals appear to be similar entities. See also Carston (2002: 56–64); Soames (2009a).

²“To give the meaning of an expression (in the sense in which I am using the word) is to give *general directions* for its use to refer to or mention particular objects or persons; to give the meaning of a sentence is to give *general directions* for its use in making true or false assertions” (Strawson 1950: 9).

(61) She is a philosopher,

the associated blueprint would contain the act type of assertion and a proposition type of the form Fx . The point is, then, that uttering a token of (61) is the normal and optimal way, in English, and in appropriate contexts, to perform an act of *asserting that a specific individual has the property of being a philosopher*. In Neale's (2005: 189) analogy, propositional blueprints are like model airplane kits; you get the manual, glue, and pieces to assemble—the construction itself is your job. The speaker's task is to make sure the blueprint furnishes the hearer with good evidence for the underlying communicative intention. The hearer's task is to reconstruct the intention on the basis of the blueprint.

A singular term e semantically encodes the instructions: *look for the salient individual or object to which the speaker intends to refer with e on the occasion of utterance*. This is of course quite minimal, but, in addition to these instructions, different kinds of singular term encode different perspectives on the intended referent. The second person singular 'you' encodes the *addressee* perspective, 'this'/'that' and 'here'/'there' seem to encode relative perspectives of abstract or concrete distances. Some indexicals, such as 'it' and 'he,' might be completely aperspectival, but this is a matter of current debate.³ Refinements aside, and ignoring tense, the blueprint semantically associated with (61) might be represented like this:

(62) $\langle \vdash_A, x_{\text{sing., fem.}}(\text{PHILOSOPHER}) \rangle$

In short, S utters (61) to instruct H that S is asserting that a unique, female individual has the property of being a philosopher. Assume o is the individual intended. Then the proposition primarily expressed by uttering (61) is not something like (62), but a singular proposition with an o -dependent truth condition.⁴ Furthermore, uttering a sentence containing a singular term communicates, by default, a presumption that the reference is edenic. When such a presumption is defeated the act of reference is revealed as defective.

Acts of coreference bring interesting complications. The cognizance constraint from Chapter 4 said, roughly, that competent, edenic speakers, in uttering a sentence containing two

³See Neale (2007: 335–345, 2008: 383–384); Rothschild & Segal (2009: 486–487).

⁴I am also ignoring the possibility that the property or concept intended is not identical but somehow related to the property encoded by $[_{VP} \text{ is a philosopher}]$. The concept intended might always be *ad hoc* relative to the VP.

singular terms e_1 and e_2 , must either intend that they corefer or that they do not corefer—but, of course, usually without thinking of what they are doing in any such theory-laden terms. Let us call this disjunctive intention the ‘identity’ intention. When the speaker is unsure or wonders explicitly whether e_1 and e_2 corefer or not, they still intend that: the terms corefer or they do not.⁵ Its presence is easily confirmed by asking speakers, post-utterance, whether what e_1 stands for is supposed to be the same as what e_2 stands for.⁶ If the speaker is indifferent, they are either playing or they lack pragmatic competence, momentarily or persistently. At any rate, they violate the constraint if they really do not have the identity intention. Uttering a sentence containing more than one singular term communicates a presumption that the speaker has the identity intention.

Now we have enough on the table to state the thesis of this Chapter. The idea is that blueprints for coreferential constructions in natural language can always encode the instruction: *look for a single object*. In such a case, when the object is found, understanding is achieved by the hearer latching on to the property intended and applying it to the object. This results in a minimal monadic proposition, even in cases where the expression uttered is polyadic.

It is well documented that human cognition in general, and linguistic interpretation in particular, are governed by a principle of minimum effort. As Daniel Kahneman puts it, “if there are several ways of achieving the same goal, people will eventually gravitate to the least demanding course of action” (2011: 35). In pragmatics, relevance theorists argue, in a similar vein, that human cognition tends automatically to maximize relevance, where the the relevance of an input for an individual is a measure of the extent to which it (i) yields positive cognitive effects and (ii) does so without requiring much mental effort (Sperber & Wilson 2005: 6).

In this spirit, I argue that human cognition is fundamentally monadic and that even if a sentence is, at surface form, polyadic and coreferential, it may really express a propositional

⁵That is to say, for sentences of the form ‘I wonder whether x is identical to y ’, the identity intention will take scope over the disjunction. Note, also, that the constraint only applies within a single clause or sentence—speakers can easily be referentially indifferent across distinct sentences or whole utterances without thereby begin pragmatically incompetent. See §4.2 above for discussion.

⁶The question is supposed to be equally appropriate for explicitly reflexive constructions and for sentences containing two proper names, even if the former is not really referential, but involves an expression bound by an antecedent referring expression.

form or blueprint that is monadic. Repeated elements are mere noise in the linguistic medium, to be ignored at the level of cognition.⁷

Monadic thesis

A propositional representation of a single object *o* can be monadic, even if the corresponding linguistic clause or sentence happens to be polyadic with respect to *o*.

The modal force of the ‘can’ here is not mere metaphysical or logical possibility, but physical possibility. I argue that this can and does happen in the actual world. Note that the thesis is relatively uncontroversial if only applied to explicitly reflexive constructions containing expressions like ‘herself’ or the Icelandic ‘sig.’ For it is agreed that in such cases the speaker may intend a monadic property. Thus there is an interpretation of ‘Joe loves himself and Jill does too,’ on which the latter conjunct picks up on the monadic property of loving oneself, expressed in the former conjunct, and applies it to Jill.⁸ But this only indicates the fact that bound reflexives and other bound pronouns are not ‘referential’ strictly speaking. More surprisingly, the monadic thesis is also intended to cover, say, sentences where the identity sign is flanked by two proper names, e.g. ‘Cicero = Tully.’⁹ This can be stated more precisely as follows. When a speaker *S* utters a sentence σ of the form $\lceil Rxy \rceil$ and believes truly *that* $x = y$ the proposition intended can be faithfully represented as $\lambda x \langle Rxx \rangle e$, where *e* is one of the singular terms properly contained in σ . For example, if I utter,

(63) Joe killed Joe,¹⁰

⁷The proposal is in some ways similar, at least in its basic motivation, to Paul Pietroski’s ‘conjunctivist’ theory of semantic composition (2005, 2007, 2011). He suggests that when concepts are lexicalized they are linked to a related monadic concept and that semantic composition proceeds simply by conjunction of the corresponding monadic predicates taking *events* as arguments. He argues that this fits well with assumptions in the minimalist program in generative linguistics, since the meaning of any expression is understood as “instructions to assemble number-neutral concepts that are monadic and conjunctive” (2011: 472).

⁸Cf. Salmon (1986b, 1992); Soames (1994). See §5.2.2 below for discussion.

⁹Thus, in *Naming and necessity*, Kripke says that it is just a confusion of philosophers to think that when you “say that Cicero is Tully, you can’t really be saying of the object which is both Cicero and Tully that it is identical with itself” (1980: 107).

¹⁰When giving examples like this in talks someone from the audience usually claims that the sentence is ungrammatical. This is not true and identity statements should be enough to establish this. I discuss the point further on p. 138 below.

having in mind only a single ‘Joe,’ what I am saying is *that the property of having killed oneself applies to Joe*. More precisely,

(64) $\lambda x \langle x \text{ killed } x \rangle \text{Joe}$

And it will always be possible to use an explicitly monadic natural language equivalent, in this case something like ‘Joe committed suicide.’ If no such rough equivalent exists it can be stipulated into existence. This, as well as (64), captures the monadic content of edenic and cognizant utterances of (63), despite its polyadic structure. It should follow, also, that (64) gives the optimal or proprietary content of (63) on this occasion of utterance.¹¹ Or so I argue.

5.2 Arguing for the monadic thesis

Suppose we discover a community of people that has remained isolated for hundreds of years. Yet they seem to speak perfect English (call their language L_1). But there is one interesting distinction between our language (L_2) and theirs. In L_1 there is no monadic predicate corresponding to ‘ x eats’ in L_2 , transitive or intransitive. As it turns out, members of the hidden community never really eat by themselves, they only feed others or accept offers of being fed by others.¹² The Feeders have a homophone of the transitive verb ‘feed’ in L_2 that serves a similar function. Instead of ‘ A eats’ L_1 contains sentences like ‘ B feeds A ’ or ‘ A is fed by B ’

Zoe the brave anthropologist visits the Feeders to study their culture. When dinner is served she makes the mistake of *eating*, in the normal way. Call this situation S , fully depicted by this drawing:

¹¹Sentence σ has p as its optimal content, on an occasion of utterance, relative to group G of speakers if, and only if, an edenic and cognizant member S of G who utters σ on that occasion, is most likely to achieve success in communication if S thereby intends to communicate p . See Chapter 4 above and Grice (1982: 299) for discussion of this notion of optimality.

¹²This could be for a number of reasons. They may think it is extremely selfish to put food in one’s own mouth, or they have somehow mutated and are unable to learn how to coordinate their hand movements with the location of the mouth—the fork always ends up in their eyes. Nowadays, they do not even remember or think about the obsolete act of eating.

Situation *S*

The Feeders are able to have a visual perceptual experience of *S* and judge *that S is the case*. Some would argue that the very content of the Feeders' perceptual experience is a proposition of this kind.¹³ We don't need such a strong assumption. Let us simply assume that the Feeders will judge or understand *that S is the case* on the basis of a nonconceptual perceptual experience of *S*. There is, at least, some propositional attitude that the Feeders bear to some proposition representing *S*. I will assume here, for sake of argument, that the Feeders *understand* what Zoe is doing in *S*, i.e. that she is consuming food, *before* they represent this fact in their public language, L_1 . L_2 -speakers also represent *S* and understand that Zoe is consuming food in *S*.¹⁴ Thus there is a property *F* such that, in judging *S* to be the case, both groups attribute *F* to Zoe.

There will, however, be clear differences in how Feeders and Eaters give expression to what they judge to be the case. The Feeders may shout out, among other things,

(65) Zoe feeds *Zoe*!

(66) Zoe feeds *herself*!

The Eaters will prefer to describe the same situation by uttering (67), without shouting in excitement.

(67) Zoe eats.

¹³Byrne (2005); McDowell (1994/1996); Searle (1983); Siegel (2012); Peacocke (1983); Thau (2002); Tye (1995). For the opposing view, see Brewer (2011); Burge (2010); Campbell (2002); Crane (2009); Gauker (2012); Travis (2004).

¹⁴Those who tenaciously believe that public languages either *are* languages of thought or that there is a perfect isomorphism between the two will clearly take issue with the assumption. I will not attempt an argument against such a position here, as it would clearly require a dissertation on its own. But I do think the argument from underspecification (§3.2) and the argument from *ad hoc* concepts (§5.2.4, below) are suggestive as to why the views in question are untenable.

Still it seems unnecessary to suppose that L_1 -speakers' cognitive representation of S is different from L_2 -speakers' representations of S in virtue of linguistic differences. In particular, the L_1 -speaker does not need to cognitively represent two individuals who happen both to be Zoe in order to represent her as consuming food: the repetition is due only to linguistic trappings. The L_1 -speaker represents Zoe as having a certain property F , it just so happens that (65) may be the most natural way for her to express this fact.¹⁵

5.2.1 Argument from the arbitrariness of language

The first argument for the monadic thesis emerges naturally from our story. As noted, it is a mere artifact of linguistic convention that (65) could be the most natural way for Feeders to express what we would express with (67). Thus one is not compelled to posit corresponding differences at the level of cognitive representation or propositional blueprints. The manner in which the Feeders cognitively represent Zoe upon observing her consuming food in this, to them, new manner, need not be structured in the same way as the sentence itself.

Here are two more specific ways to flesh this out. First, suppose we eliminate all Rxx -constructions for some equivalent Fx -construction to make a new language. If needed, the latter can be a neologism and the equivalence a matter of stipulation of the Kripkean 'schm'-variety (Kripke 1980: 108). Intuitively, this would have zero effect on the expressibility of the language. Nothing of importance is lost with the Rxx -constructions if the monadic replacements really are equivalent. Admittedly, this point begs the question against naïvely Fregean theories of content. A certain specific mode of presentation will be lost, namely the mode associated with Rxx -type expressions. And if propositions only contain such modes of presentation the envisaged elimination makes for a loss in expressibility. But this is not an attractive feature

¹⁵According to current syntactic theory a sentence like (65) is *noncoindexed* since, by Principle C in Binding Theory, R-expressions must be free in all categories. So the sentence would be represented as, 'Zoe₁ feeds Zoe₂,' to indicate that it is neither coreferential nor noncoreferential *in virtue of linguistic meaning or grammatical structure* (Fiengo & May 1994). This may appear incompatible with the monadic thesis but it is not. Binding Theory, of course, allows that noncoindexed expressions can be covalued and coreferential. My point in this Chapter is that, given certain independently motivated idealizations, coreferential utterances of noncoindexed sentences can be uttered to express monadic contents, where the associated blueprint indicates that this is so.

of Fregeanism anyway, for it assumes that any trivial addition to a language would add to its expressibility. So, if we stipulated *triadic* constructions into existence, equivalent to the original dyadic one, this would give us yet another mode of presenting the relevant proposition. But new beliefs are not produced by mere ‘stroke of the pen,’ in Grice’s phrase (Grice 1969b: 140, cf. Evans 1982: 50).

The elimination itself can be conceived of in many different ways. The new language could systematically distinguish between relations to oneself and relations to others by using mandatory prefixes like ‘auto-’ and ‘allo-.’ ‘Eating’ then becomes ‘autofeeding,’ ‘feeding’ becomes ‘allofeeding’ and there would be no such thing as $\lceil x \text{ feeding } x \rceil$. Another way would be to stipulate that every relation between a thing and itself automatically gives rise to a distinct lexicalized concept. Thus we will always have $\lceil x \text{ eats} \rceil$ instead of $\lceil x \text{ feeds } x \rceil$ and $\lceil x \text{ committed suicide} \rceil$ instead of $\lceil x \text{ killed } x \rceil$. Arguably, neither option decreases the expressibility of the language.

Secondly, suppose that for all existing *Rxx/Fx*-equivalences in the language we eliminate the *Fx*-constructions. Plausibly, this would not eliminate ability to entertain and communicate monadic contents for which the speakers only have *Rxx*-constructions to represent linguistically. Take $\lceil x \text{ committed suicide} \rceil$ and the strictly reflexive use of $\lceil x \text{ killed } x \rceil$ as these occur in English. One thing seems clear. Even if one’s public language only contained the latter, one would still be able nonlinguistically to entertain a monadic proposition to the effect that someone is a self-killer.¹⁶ And a perspicuous way to represent this ability is to say that the thinker entertains a thought the content of which is a monadic proposition, in this case something like: $\lambda x \langle x \text{ killed } x \rangle_e$.¹⁷

The first elimination begs the question against Fregeans. This one does so against theorists who believe (i) that there is a language of thought (LOT) and, (ii) the language of thought is the natural language idiolect of the speaker. I am happy to concede the LOT hypothesis but

¹⁶The argument is intended to mirror Kripke’s well-known claim that speaking an explicitly descriptivist language would not make it impossible for one to entertain and communicate object-dependent contents (1977: 265–266). See also Evans (1981: 321), who makes a similar point with respect to indexicals.

¹⁷Or, in terms similar to Pietroski’s (2005), the thinker entertains a thought about an event of self-killing and thinks of some person as having the Thematic Role of the Agent of that event.

(ii) is a different matter. Here I only note the incompatibility as I do not have a knock-down argument against (ii).¹⁸

Still, some may ask where this leaves confused and coreferentially indifferent speakers. Now I have argued extensively that these can be put to one side, but let's consider the question in the current context. Does the elimination of $\lceil x \text{ eats} \rceil$ or any monadic equivalent lead to the loss of expressibility for such non-ideal speakers?

Consider the utterance,

(68) Zoe feeds her,

as uttered by the Feeders' Manager of Feeding Partnerships (MFP) who points to Zoe and intends to refer to Zoe with 'her.' But the MFP does not recognize Zoe, having just recently picked up the name from friends. Doing her job, the MFP is simply pointing to the people in the dining hall and assigning each a feeding partner, pretty much at random. She does not realize that she appears to be referring to the same person twice. Indeed, being a Feeder, the MFP would *never* assign feeding and being fed to one and the same person. It is not clear whether she could even have entertained the possibility before seeing Zoe eat.

Now, what is the proposition expressed here? I am not sure this question has an interesting answer. The MFP intends to express a proposition about two individuals but fails to do so. Does she then express a proposition about a single individual? As I have argued, there is not much theoretical point in talking about propositions expressed unintentionally. The hearer (*H*) may either be (i) aware of the MFP's confusion or (ii) be confused as well. In case (i) *H* will be able to find the 'right' interpretation: *H* will really know what the speaker intended and why this intention could not be fulfilled or was incoherent. In (ii) *H* will think that Zoe is not the referent of 'her,' and believe that the MFP succeeded in expressing a proposition about two distinct individuals. But if the proposition really only contains a single individual, Zoe, then the proposition that the interpreter thinks they have grasped does not exist. So they merely thought they were grasping such a proposition. Furthermore, when the MFP realizes that she

¹⁸See again page 87 in Chapter 3 above, where I conceded that Mentalese may partly be couched in a public, natural language. See Devitt & Sterelny (1999: 138–146); Fodor (1975, 2008); Harman (1973). As Fodor (1998: 9) notes the representational theory of mind, as such, "tolerates the metaphysical possibility of mental representation without thought."

appeared to *H* to refer to Zoe twice, she will not think that she actually intended to express the proposition *that Zoe feeds herself*. The MFP intended no such thing: so it makes no sense to say that this is what she really meant. The argument here is similar to the ones presented in Chapter 2 and, in like manner, shows that the importance of confusion puzzles has been greatly exaggerated in the philosophy of language.

5.2.2 Argument from reflexivity

If ‘herself’ in (66) (*‘Zoe feeds herself!’*) is such that it (i) represents Zoe in some sense, (ii) is bound by ‘Zoe’ in (66) and (iii) (66) can be uttered to express a monadic proposition, the monadic thesis must be true, at least in this weak sense. There are two linguistic expressions in (66) that serve to represent Zoe, but there clearly is a reading of (66) on which a monadic (reflexive) property is applied to Zoe (Salmon 1986b, 1992; Soames 1994). This is the by now familiar reflexive singular proposition that can be represented as,

(69) $\lambda x \langle x \text{ feeds } x \rangle \text{Zoe}$.

It has been established, however, that explicit reflexives can also have ‘strict’ readings, so that the property applied to Zoe would be, $\lambda x \langle x \text{ feeds } \text{Zoe} \rangle$. The reading given in (69) is the ‘sloppy’ reading. Traditionally, the strict/sloppy distinction was only thought to apply to nonreflexive pronouns and reflexives were believed to be consistently sloppy (Partee & Bach 1984; Williams 1977). But as Arild Hestvik (1995) showed, VP ellipsis data, especially when elision occurs in a subordinate clause, brings out the distinction for reflexives as well.

(70) Bill shaved himself better than Becki did.¹⁹

The sentence has at least two readings, represented somewhat tortuously here:

(71) $[\lambda x \langle x \text{ shaved } x \rangle \text{Bill}] \text{ better than } [\lambda x \langle x \text{ shaved } x \rangle \text{Becki}]$

(72) $[\lambda x \langle x \text{ shaved Bill} \rangle \text{Bill}] \text{ better than } [\lambda x \langle x \text{ shaved Bill} \rangle \text{Becki}]$

Thus (71) ‘sloppily’ applies the property of *having shaved oneself* to both Bill and Becki, while (72) does the same to the ‘strict’ property of *having shaved Bill*.

¹⁹This is based on Hestvik’s (ibid., 213) example of what he calls the ‘subordination effect.’

My argument, here, is that it is sometimes possible to extend sloppy readings to sentences containing only proper names. Thus even something with the explicitly dyadic form of (65) ('Zoe feeds Zoe!') can be intended 'reflexively' by the speaker such that a monadic property is being applied to Zoe. This cannot, however, be established simply with VP-ellipsis data as in the case of (70). But other kinds of contexts can be constructed to support the view that an utterance of something like (65) can have monadic equivalents or, even, be faithfully reported via monadic constructions. Showing this would clearly be evidence for the monadic thesis.

Accordingly, let us consider two kinds of examples. First, a context in which the dyadic and reflexive forms are equally appropriate. Second, speech act reports where the reported sentence is dyadic but the report is explicitly reflexive.

It is sometimes suggested—when I present this material at conferences for example, but also in many writings by linguists—that sentences like (65) are ungrammatical. This is clearly wrong.²⁰ Simply imagine someone who is explaining a new game to a group of Feeders where the participants are supposed to *eat*, just like Zoe did. The instructor says: 'Everybody is supposed to feed!' And someone asks: 'Each other or ourselves?' To which the instructor replies: 'Rachel feeds Rachel. Joey feeds Joey, and so on. *You do the same*. Get it?' In this case doing the same is *not* to feed Rachel or Joey, but to feed oneself and the explicitly reflexive form 'x feeds herself' would have served the same purpose. In this context the speaker's reference is both edenic and cognizant, and the audience, it is safe to say, will understand perfectly what is being said. Thus uttering an explicitly dyadic sentence, where a relation is flanked by two names, can bring to salience the sloppy reflexive property rather than the strict one.

Not to overstate the point, but it is absolutely clear that reflexive and pronominal constructions indicate the monadic interpretation much more *strongly* than dyadic constructions with proper names do. Thus in both 'Janice loves her mother, Becki does too' and 'Bill loves himself, so does Becki' there is a strict/sloppy ambiguity. It appears absurd to suggest the same

²⁰Writing about Icelandic syntax, Höskuldur Þráinsson (2005: 519) says that a sentence like 'Jose shaved Jose' cannot be understood except if there are two Joses at issue. Apart from the point made in the text, it seems like homophonic confusion puzzles like Kripke's Paderewski case require sentences like 'Paderewski is Paderewski' to be, at the very least, *grammatical*. My argument has been that such confused utterances are *pragmatically* defective. See page 21 in Chapter 1.

ambiguity applies to, say, ‘Bill loves Bill, so does Becki’ or ‘Cicero is Cicero, so is Tully.’ Here only the strict reading is possible in the second clause (more on this point below).

Secondly, consider speech act reports. Joe is a small time crook planning to shoot and kill a man called Chuck. Botching the job, he shoots himself instead. Two friends of Joe’s, call them *A* and *B*, are talking. Both knew about his original plan but only *A* knows what really happened. So, when *A* brings this up, *B* *expects* to hear something about two distinct individuals, one of which killed the other. Thus *B* misunderstands *A* at first.

A: ‘Joe killed himself.’

B: ‘What? Joe killed whom?’

A: ‘Listen, Joe killed *Joe*.’²¹

In reporting what *A* said by uttering ‘Joe killed Joe’ on this occasion, somebody might perfectly well utter: ‘*A* said to *B* that Joe killed himself,’ or, even, ‘... that Joe committed suicide.’ Admittedly, speech report data is by no means conclusive, but, as Elisabeth Camp argues,

[...] our ordinary practice of speech-reporting is sensitive to a certain standard of explicitness, and [...] speakers can legitimately object to reports that disregard this standard. Specifically, it is normally only appropriate to report speakers as having ‘said’ contents to which they have *openly and obviously* committed themselves by their utterance. (2006: 286, emphasis in original)

And, surely, it is hard to conceive of a standard of explicitness such that edenic, cognizant utterances of *Rxx*-constructions cannot be reported with an equivalent *Fx*-construction. People who murder themselves usually commit suicide at the same time.

Gary Ostertag (pers. comm.) raises the following objection to the argument from reflexivity. Even if it is conceded that *Rxx-Fx* pairs *can* be expressively equivalent, why is it that we never get a sloppy reading of sentences like (73)?

²¹When testing this kind of case on people, they sometimes say they can’t hear the last utterance as intended to refer to a single Joe, the speaker must be referring to two Joes. This reaction is cancelled, however, by adding that the later occurrence of ‘Joe’ is accompanied by *A*’s pointing to a picture of Joe.

(73) Zoe feeds Zoe. So does Bill.

And, without doubt, by uttering the second clause in (73) the speaker will intend *that Bill feeds Zoe*, not *that Bill feeds himself*. This is, surely, *prima facie* evidence against the monadic thesis. Note, however, that such a negative conclusion only shows that this sort of VP-ellipsis test does *not* predict a strict/sloppy ambiguity for (65). A similar example, using a coordinating conjunction, would suggest the same negative conclusion for reflexives. Yet examples like Hestvik's, such as (70) above, indicate otherwise. Clearly, the two examples on display here are not as dispositive as Hestvik's but they definitely provide support for the monadic thesis.²²

5.2.3 Argument from middle voice

Modern Icelandic, like Ancient-Greek and a few other languages, has one voice in addition to the passive and the active. This is called the 'middle' or 'mediopassive' voice.²³

In many cases the Icelandic language allows speakers to use a simple syntactic operation to transform regular transitive verbs into intransitive reflexive verbs. The operation consists in adding an *-st*-suffix to the infinitive.

(74) Zoe¹ matar sig₁.
Zoe feeds REFL
'Zoe feeds herself.'

(75) Zoe matast.
Zoe feed-*st*-SUFFIX
'Zoe eats.'

(76) Zoe borðar.
Zoe eats
'Zoe eats.'

²²Ostertag adds that there is no way to evaluate this response if we don't also have some general way of determining when $\lceil Rxx \rceil$ -type utterances are really the expression of $\lceil Fx \rceil$ -type propositions. True, I do not have a criterion by which to categorize every utterance but this is hardly the be-all and end-all of the current proposal. My aim is only to show that, given certain specific idealizations, there are cases where this is indeed how one ought to describe what the speaker expresses. See, however, the discussion about how intentionalism describes constituent deletion or free impoverishment in §5.3 below, where the monadic thesis is explained in terms of communicative purposes.

²³See Thráinsson (2007: 283–293) for an overview.

In this particular case, (75) is synchronically derived from (74) as the *-st*-suffix is a truncated form of the non-possessive reflexive ‘sig.’ Admittedly, this description is somewhat controversial because there is a lot of semantic and syntactic variation in the class of *st*-verbs (cf. Anderson 1990). Nonetheless, it is clear that there is a large subclass of such verbs where the reflexive and the middle voice are straightforwardly equivalent in meaning. Strictly speaking, however, (74) and (75) can’t be used to show this, since (74) is quite artificial and would never be used by competent speakers. There are, however, lots of such reflexive-middle voice pairs whose use is simply the same: ‘Zoe leggur sig’ and ‘Zoe leggst,’ both have roughly the meaning of ‘Zoe lays herself down.’ ‘Klæða sig’ and ‘klæðast,’ both have the meaning of ‘to dress oneself,’ and so on. I’ll pretend (74) belongs to this class.

More tellingly, however, the middle voice is taught to native speakers, like the passive, as a simple derivation of the active voice. Elementary grammar books will say things like: the passive is derived from the active by (i) adding ‘to be’ as an auxiliary, (ii) putting the main verb into its past participle form, (iii) making the object the subject of the sentence, etc. And the instructions for creating the middle voice simply say: suffix ‘st’ to the infinitive of the main verb. In many grammar books I have checked, the equivalence of the reflexive and middle voice forms is noted for the benefit of the student.

Now, (75) is strictly synonymous with (76). Thus we have here an actual natural language which supports the principle driving the arguments given before, namely that an explicitly coreferential and reflexive construction can have monadic equivalents. The structure of (74) is *Rxx* but it is still strictly equivalent to an expression with an *Fx* type structure. This is evidence for the monadic thesis, but only if it is conceded that there are contexts in which (74) can be uttered to express the same proposition when ‘Zoe’ is substituted for ‘sig/herself.’

5.2.4 Argument from *ad hoc* concepts

When one watches the final scene in the Hungarian film *Taxidermia* and cognitively represents a taxidermist stuffing and mounting his own body, one’s representation is different in kind from one in which a taxidermist prepares to stuff and mount, say, the skin of a lion. One observes that the taxidermist has a certain property *F* but cannot find a more suitable linguistic expression for

this than the dyadic predicate $\lceil x \text{ stuffs } x \rceil$. Consider also irreflexive relations. Suppose—if that is even possible—we discover an object x such that x is higher than x . Our cognitive situation is then similar to the Feeders': we represent an object as having a property but the present state of our public language is such that we cannot think of a better way to communicate this fact than by using a dyadic predicate.²⁴ Perhaps M.C. Escher's drawing 'Ascending and descending' provides an example of a point in space x such that x is higher than x .²⁵

Here I will argue that, at least in some cases, when an irreflexive relation is applied reflexively for the first time a new concept is expressed. The new concept is monadic and can be explicated in terms of *ad hoc* concepts. This notion comes from Barsalou's (1983, 1987, 1999) work on conceptual structure and has been used in the theory of meaning and communication by Sperber & Wilson (1998) and others.²⁶ Sperber & Wilson propose a minimal characterization of concepts which is neutral between different theories of their actual structure: "A concept [...] is an enduring elementary mental structure, which is capable of playing different discriminatory or inferential roles on different occasions in an individual's mental life" (1998: 35). A mere difference in perceptual discrimination—different shades of color, for instance—does not necessarily correspond to a conceptual difference. And momentary representations of individuals do not, automatically at least, give rise to new concepts (e.g. *that tree I saw*). Despite this many concepts are *ad hoc* in the sense that they are constructed and communicated 'on the fly' without being lexicalized in the speaker's public language. These concepts are enduring mental structures since they are stored in long-term memory and can, with due time, become lexicalized.

One of Wilson & Sperber's (2002b: §3.6) stock examples is as follows. Two people, *A* and *B*, are planning a biking trip. *A* feels unfit and *B* knows this. *B* utters:

(77) We could go to Holland. Holland is flat.

Their claim is that on this occasion *B* expresses an *ad hoc* concept of flatness by uttering the

²⁴Note, however, that the thought expressed by $\neg\exists x(x \text{ is higher than } x)$ is perfectly coherent and well within our present conception. But this is necessarily a *general, quantificational* thought. The point here only applies to thoughts of particulars.

²⁵Graham Priest (2006: 59–60) argues that this is so. I am not endorsing his argument here.

²⁶Cf. Carston (2002, ch. 5); Prinz (2002, ch. 6).

word ‘flat.’ This concept is not lexicalized in English as spoken by *A* and *B*. The concept can probably be encoded or approximated by constructing a more complex expression, such as ‘ $\lceil x$ is a sufficiently flat country to satisfy your current biking preferences \rceil ’.²⁷ But it is clear enough that the concept is not encoded by the expression ‘flat.’ ‘Flat’ just happens to serve as the most economical way for *B* to indicate the *ad hoc* concept in question on this occasion of utterance. Further, this concept surely satisfies the minimal characterization since *A* and *B* could want to have recourse to this expression-concept pair more frequently in the future. The practice could replicate itself and create a need for a distinct lexical item in the public language.

To be concise, an *ad hoc* concept C^* can be defined in terms of the present state of a public language *PL*: At *t*, *PL* is such that C^* is not lexicalized and the most efficient way to express C^* in context is to utter a simple expression that encodes a related concept *C* (such that C^* implies *C* or *C* implies C^* or ...).

Here is how the Feeder’s situation is similar to the biking situation. When a Feeder observes *S* and utters (65) she thereby expresses a feeding concept C^* which is *ad hoc* relative to the state of the public language. It is *ad hoc* because C^* can relate an object to itself and is roughly equivalent to the monadic concept of eating. The Feeder actually expresses a proposition in which C^* is applied to Zoe. The Feeders had never expressed C^* before or mentally represented a situation in which C^* is instantiated. It is clear that if the Feeders start attaching significance to eating (or ‘auto-feeding’ to be more precise) as distinct from feeding, they will be motivated to lexicalize C^* . Thus, it can be a matter of mere linguistic convention that speakers relate things to themselves. Such a manner of speaking is always equivalent to some monadic construction.

Perhaps monadic forms, when a salient and equivalent dyadic form exists, are reserved for cases where people think some significant distinction is being marked. Committing suicide is roughly the same as murdering oneself. And murdering oneself is different enough from murdering another that the concept is lexically distinct. This just means that the concept of

²⁷This point is denied by Wilson & Sperber, but, as far as I can tell, without argument. They state that this *ad hoc* concept of flatness is “neither encoded nor encodable in English as spoken by [*A* and *B*] at the time of their exchange” (ibid., 76). But just as the argument for underspecification should not depend on there being *no such thing* as an eternal sentence (cf. §3.2.5 above), there is no need to deny that a concept is more or less *encodable* to show that there are *ad hoc* concepts.

suicide is an enduring elementary mental structure that plays different roles in an individual's mental life on different occasions. It is still the case that when Jill murders Bill and then commits suicide, Jill stands in the same relation to Bill as she does to herself, namely the murdering relation.

But there are counterexamples. It may seem like the concepts-pairs in question are not strictly equivalent.²⁸ So, for instance, self-feeding is not equivalent to eating, for one can eat by being fed. Similarly, if Joe killed himself purely by accident—the gun backfired when he shot at Chuck—he didn't really *commit suicide*. The concepts are clearly not equivalent in extension. It may be true, however, that he *accidentally* committed suicide.

My argument circumvents the objection, however. When the feeding relation relates a single individual to itself there is always some monadic property F such that the individual must also instantiate F . In this case F is very similar to the property of eating, although they are not strictly equivalent. Thus, the monadic thesis only involves the claim that any reflexive instantiation of a dyadic relation entails that *some* extensionally equivalent monadic property is also instantiated. And, conveniently, this monadic property can be purely stipulative and *ad hoc* relative to the speaker's idiolect or linguistic community.

A related objection was brought up by Gary Ostertag (pers. comm.). He worries that the monadic thesis is at variance with the validity of certain unassailable inferences. So, for example, is it the case that only some instances of $Ga, Fa \models (Ga \wedge Fa)$ are valid because the conclusion may or may not contain a twice over? No, the schema is definitely valid. Remember that the monadic thesis only applies at the level of individual clauses or sentences and I construe Ga and Fa as representing two distinct clauses conjoined in the conclusion. The monadic thesis is thus restricted because it is grounded in the cognizance constraint, which is restricted in the same way, and for very good reason. As I argued in Chapter 4 the constraint is quite compelling at the level of the individual clause but not so across distinct singular term clauses uttered on different occasions.

²⁸Thanks to Jesse Prinz from pressing this point.

5.3 Mismatch between meaning and content

I assume, from now on, that the monadic thesis is probably true: edenic and cognizant utterances of a coreferential constructions can indeed be associated with monadic propositional forms. Note, however, that the thesis and the attendant examples are much more radical than the ‘flat’-case just considered. The latter conforms to traditional examples of underspecification and unarticulated constituents in the literature in that it does not entail a structural mismatch: it merely requires a mapping from the concept of flatness to the *ad hoc* concept of relevant-bicycle-flatness. The input/output pairs to this mapping will have the same representational structure. The concept is *articulated* by the expression ‘flat’ but is still *underspecified* by the linguistic meaning of ‘flat.’

In line with the argument for underspecification, i.e. thesis (T) in Chapter 3, let us assume that a speaker can utter,

(78) It’s raining,

on a given occasion, and thereby express the proposition that it is raining *in New York*. This occurs without there being any overt expression in (78) with which the speaker refers to New York, making the city an unarticulated constituent of the proposition expressed. Some theorists would describe this as a case of ‘free enrichment,’ meaning that the addition of a location is not linguistically mandated (cf. Recanati 2010).

As in the case of *ad hoc* concepts, this description does not imply the rejection of the common sense idea that the structure of natural language representations maps comfortably onto the structure of cognitive representations: that is to say, the assumption that ‘sentences’ in the language of thought—whatever they are—correspond to structurally similar sentences in the speaker’s idiolect. All that is needed is (i) the addition of one extra propositional constituent, i.e. New York, and (ii) this constituent, undeniably, could have been added explicitly by making a corresponding addition at the level of linguistic representation. The speaker could have uttered (79) instead.

(79) It’s raining in New York.

Consider, briefly, two popular theories of why (78) can be uttered to express something about New York. First, there is hidden indexicalism (cf. Stanley 2007). On this view, there is a hidden indexical expression in (78) at the level of logical form. In essence, then, (78) is not different in kind from overtly indexical sentences. There is an aphonic location variable attached to [_{NP} raining]. The variable gets New York assigned as semantic value in the context of utterance. Thus the gap between (78) and the proposition expressed by uttering (78) on an occasion is nothing extraordinary.

Secondly, there is minimalism. The sentence in (78) expresses a minimal proposition that is determined by the conventions of the language to which (78) belongs (Borg 2004, 2012; Cappelen & Lepore 2005). This proposition has nothing to do with New York or any other location: it is the content which every utterance of the sentence type (78) shares in common with every other utterance of (78), i.e. *that it's raining* (call this proposition *p*).²⁹ A minimalist then explains the fact that (78) may seem elliptical for (79) in certain contexts by adding that a speaker may indirectly *communicate* the proposition literally expressed by (79), call it *q*, by directly expressing *p*. In such contexts, the speaker assumes that the location intended is obvious enough for the hearer to realize, maybe by processing *p* first, that *q* is what they primarily want to communicate.

The Feeders' case is not adequately described in any of these familiar terms. Or so I argue. The monadic thesis requires 'deletion' or 'free impoverishment'—where relations are modified and argument positions removed—at the level of cognitive representation while unarticulated constituents require 'additions.' By assumption, the monadic proposition expressed by the Feeder contains Zoe only once although the sentence uttered contains 'Zoe' twice. And it is not true that the same proposition can be expressed by uttering a sentence where one of the occurrences of 'Zoe' has simply been deleted. This would result in ungrammaticality or change in meaning. Thus, if anything is deleted in any sense, it can only be deleted at the level of propositional representation.

Let us look at the theories themselves. First, positing a hidden indexical to accommodate

²⁹Here the question ought to be raised whether there actually is any such proposition and, if there is, what its truth condition is. Recanati (2002, 2010: §2.1) presents an example intended to show that such a context is really possible.

the monadic thesis makes scant sense. There is no indexical such that it occurs, covertly or overtly, in a dyadic sentence σ and maps σ onto an equivalent monadic proposition. Neither is there an indexical that deletes an occurrence of a propositional constituent whenever that constituent is referred to with two or more singular terms in the same sentence.

Secondly, minimalists would reject the monadic thesis out of hand, unless the minimal proposition expressed by (65) is actually thought to be monadic in some cases. But it is highly unlikely that this could be squared with minimalist principles. The other minimalist possibility would be to say that a coreferential dyadic proposition can be expressed in order to convey another equivalent proposition with monadic structure. But since the latter proposition is equivalent the hearer has no need for it after the dyadic proposition has already been communicated and, so, the suggestion has no motivation at all.

Instead of taking of either of these theories on board, let us consider the propositional blueprint account in more detail, as it applies to the monadic thesis. Here is how the ‘deletion’ of propositional constituents is to be explained on that picture. Uttering the irreflexive dyadic predicate ‘ x feeds y ’ in the language of the Feeders communicates a presumption that the speaker intends to express a proposition containing *two* non-identical objects as the relations’ *relata*. Arguably, uttering ‘ x is higher than y ’ normally creates the same type of presumption in our own language. But the presumption can certainly be defeated in both cases.

In interpreting (65) a Feeder will probably start by treating Zoe as two different objects but, realizing that something else is really intended, they will correct the mistaken assumption and represent Zoe as having some property F roughly coextensive with the property of eating. The Feeder may recognize that the content of this representation is equivalent to the content of the perceptually grounded judgment *that S is the case* or, even, *that Zoe was the agent and patient of the feeding-event*. Thus the hearer grasps the fact that both occurrences of ‘Zoe’ in (65) refer to the same individual and thus represents her as having some observed property. This kind of story does not require the interpreter to reason from a reflexive dyadic proposition to a monadic proposition and enables the speaker to intend to express the monadic proposition directly by uttering (65).

It helps, I believe, to take the airplane model analogy somewhat literally. When the same

object is represented more than once in a single model-blueprint this does not usually mean that one is supposed to build two such objects or build the same object twice. Normally, it is just a way of representing the same object from different angles in order to attribute some complex property to it. And this may be because it is difficult—relative to the medium of representation—to express this complex property by using only a single representation of the object in question. I have argued that (65) ('Zoe feeds Zoe'), as it occurs in the Feeders' language, should be construed in a similarly monadic way. And, by extension, utterances that are manifestly and obviously coreferential or reflexive, in our own language, can be intended to communicate contents that are in fact monadic.

But what communicative purpose could be served by freely impoverishing contents in this way? Briefly, the communicative and cognitive purpose of minimal effort. An interpreter who understands an edenic and cognizant coreferential utterance will implicitly recognize that the speaker used a circumlocution—remember the explanation of the game and the failed assassination attempt in §5.2.2 above—because it just happened to be the optimal way to get a simple message across in the context. And, if the whole argument of this dissertation is on the right track, this gives part of the 'normal conditions' for the utterances in question, in Millikan's sense. The abnormal conditions give rise to various kinds of corruptive effects on communication and cognition more generally.

5.3.1 Overarticulated constituents

John Perry (1986) originally described cases like (78) by saying that there can be thoughts without representation. More specifically, there can be constituents of propositions serving as objects of propositional attitudes, like *thinking*, such that nothing in the sentence corresponding to the proposition corresponds to that constituent. As I have shown, there are also representations without thoughts. First, in the sense that the 'thought' itself can be confused and corrupt, and this can be masked by the linguistic representation, which, by itself, suggests that there is a coherent thought in the mind of the speaker. Secondly, propositional constituents may correspond to more than one singular term in the sentence uttered. And, if the monadic thesis is correct, additional occurrences of coreferring singular terms don't

necessarily correspond to more occurrences of the constituent in the proposition.

Clearly, there are many kinds of redundancy or repetition in natural language. I have focused only on what I will call ‘structural’ repetition as distinct from ‘pragmatic’ ones. The latter often serve quite specific communicative purposes, such as emphasis or the expression of excitement. Sperber & Wilson (1986: 219) give many nice examples of the variation here. Thus, uttering ‘I went for a long, long walk’ may be intended to say that the walk was *very* long. ‘There’s a fox, a fox in the garden’ to show that the speaker is excited about the fox, and so on.³⁰

Structural repetition occurs, however, when semantic and pragmatic competence in a given natural language simply requires that the speaker makes repeated reference to the same object—let’s focus only on singular terms here—in expressing some particular proposition. The degree to which the speaker is so required varies, of course, from case to case. Anyhow, structural repetition or overarticulation comes in three broad categories. First there is simple coreference, which we have been focusing on until now, where some polyadic relation is predicated of a single object. Secondly, there is *complex* reference, which is when the singular term has coreferential singular terms as immediate constituents. Thirdly, there is the class of *appositives*.

Take each of the latter in turn. To explain complex reference, consider the following list of utterances.

- (80) a. It’s raining.
 b. It’s raining here.
 c. It’s raining here in Paris.
 d. It’s raining here in Paris, Illinois.

The proposition expressed by uttering (80a) would contain an unarticulated constituent, i.e. a certain city, which would have been articulated by uttering one of (80b)-(80d), keeping the context fixed in other respects. Call the context of utterance *C* and assume that, in *C*, (80a) would have been quite sufficient to communicate that it’s raining *here in Paris, Illinois*. This

³⁰Some theorists have argued that what I call repetition, structural or pragmatic, poses serious problems for relevance theory (Doerge 2013; Jucker 1994). I am not convinced that this is so.

implies that, in *C*, any of (80b)-(80d) would have been an overarticulation of the location of the rain. If (80a) is sufficient in *C* then, by the time the speaker *S* has finished pronouncing ‘raining,’ *S* has already made one (implicit, unarticulated) reference to Paris. By the same token, *S* will have made two references to Paris by the time *S* starts pronouncing ‘in Paris’ in (80c). And so on.

Yet the location of the rain is still *underspecified*, at the very least, by the linguistic meaning of (80c) since there are many places called ‘Paris.’ This is another reason, in addition to the point about illocutionary force in §3.2 above, to think that not all cases of underspecification involve unarticulated constituency. For, otherwise, one would expect that overarticulation would lead to overspecification, or, at least, not to underspecification.

Appositives are similar.³¹

- (81) a. Obama likes cats.
 b. Obama, the 44th president, likes cats.³²
 c. Obama, the 44th president, standing over there, likes cats.

Certainly, there are contexts in which only the most verbose sentence on either list will secure uptake—otherwise it would presumably never be uttered at all. But appositives serve many other functions as well, for example to indicate in passing some relevant fact about the individual referred to: ‘Joe, the man wearing a ridiculous bow-tie, ...’

The pragmatic-structural distinction is not precise, but it need not be. The point of the examples is to show other phenomena to which the monadic thesis, arguably, applies. These are sentences containing two or more singular terms, each of which is intended to refer to the same object, and the utterance as a whole is primarily intended to attribute a simple monadic property to that object. The added material is a secondary, non-central aspect of the act of reference. Compare this to pure redundancies, such as ‘Audible to the ear,’ ‘Re-elected to another term’ (Wallace 2006: 66). Someone might argue that the latter phrase ‘literally’ means that a person was elected for the third time in a row, but the speaker probably intended to

³¹Thanks to Stephen Schiffer who brought this to my attention.

³²For discussion about how to analyze appositive uses of definite descriptions, see Neale (1990: 116n55, 2008: 414). Searle’s (1979b: 145–147) early discussion of the difference between primary and secondary aspects under which reference is made is also relevant here.

say only that someone was *re-elected*, full stop. Normally, speakers do not even notice the redundancy.

Many cases of coreference are conventionally entrenched versions of this phenomenon, in the sense that the conventions or regularities of the language could easily be such that for every occurrence of an *Raa* construction we would have some *Fa* construction instead. My claim has been that this point applies equally to obligatorily reflexive verbs—e.g. $\lceil x \text{ perjures } x \rceil$ —and any edenic, cognizant utterance of a dyadic coreferential sentence. When there actually is coreference, the blueprint encodes the instruction: look for a single object. When the speaker intends to refer to two distinct objects, the propositional blueprint says: look for two objects. In this sense, $\lceil Rxx \rceil$ and $\lceil Rxy \rceil$ results in different sentences. As Kit Fine (2007) would put it, the variables are coordinated in the first but not in the second. The monadic thesis is similar, then, to some aspects of Fine’s semantic relationism: even proper names are ‘coordinated’ when they corefer in the same sentence. Importantly, however, I don’t think there is any such relation of coordination across different contents or different sentences. As argued in Chapter 4 the cognizance constraint only applies *within* the clause or sentence.

Conclusion

While this should not be taken as an endorsement of his theory of belief and singular thought—for it is afflicted by residual verificationism³³—my argument in this Chapter is similar to one made by Gareth Evans (1979, 1982). His point is about two sentences, one containing a definite description and the other a ‘descriptive’ name, where it has been stipulated that the name refers to the person, if there is one, who uniquely satisfies the descriptive condition: ‘Julius is *F*’ and ‘The inventor of the zip is *F*.’ Evans writes:

Someone who understand and accepts the one sentence as true gets himself into exactly the same belief state as someone who accepts the other. Belief states are distinguished by the evidence which gives rise to them, and the expectations, behaviour, and further beliefs which may be derived from them (in conjunction

³³Cf. Devitt (1985: 229); Hawthorne & Manley (2012: 78); Schiffer (1988: 38–39).

with other beliefs); and in all these respects, the belief states associated with the two sentences are indistinguishable. (1982: 50)

I think the formulation is far too strong but, still, the weaker point that belief states associated with the two sentences appear to be similar in interesting ways is undeniable. In similar fashion, edenic and cognizant utterances of a given coreferential construction will normally express a belief that is all but indistinguishable from the belief expressed, in the same context, by uttering its monadic linguistic counterpart. And this is so even if, going beyond Evan's broadly functionalist view as stated here, the intrinsic structure of the associated mental representation goes into the individuation of the belief state.

Bibliography

Page references are always to reprints when listed.

Abbott, B., 2010. *Reference*. OUP.

Ackerman, D., 1979. "Proper names, propositional attitudes and non-descriptive connotations." *Philosophical Studies*, 35(1):55–69.

Alexander, J., 2012. *Experimental philosophy: An introduction*. Polity.

Anderson, S. R., 1990. "The grammar of Icelandic verbs in *-st*." *Syntax and Semantics*, 24:235–273.

Anscombe, G. E. M., 1957. *Intention*. HUP.

Appelbaum, I., 1996. "The lack of invariance problem and the goal of speech perception." *ICSLP 1996. Proceedings of the Fourth International Conference on Spoken Language Processing*. vol. 3, pp. 1541–1544.

Audi, P., 2012. "Grounding: Toward a theory of the in-virtue-of relation." *The Journal of Philosophy*, 109(12):685–711.

Austin, J. L., 1957. "A plea for excuses." *Proceedings of the Aristotelian Society*, 57:1–30. Repr. in Austin (1979), pp. 123–152.

———, 1962. *Sense and sensibilia*. OUP.

———, 1979. *Philosophical papers*. OUP, 3 edn. First publ. in 1961.

Bach, K., 1987. *Thought and reference*. Clarendon.

———, 1994a. "Conversational implicature." *Mind & Language*, 9(2):124–162.

———, 1994b. "Semantic slack: What is said and more." S. L. Tsohatzidis (ed.), *Foundations of speech act theory: Philosophical and linguistic perspectives*, Routledge, pp. 267–291.

———, 1999. "The myth of conventional implicature." *Linguistics and Philosophy*, 22(4):327–366.

———, 2005. "Context *ex machina*." Z. G. Szabó (ed.), *Semantics versus pragmatics*, Clarendon, pp. 15–44.

Bach, K. & Harnish, R., 1979. *Linguistic communication and speech acts*. MIT Press.

Bar-On, D., 2013. "Origins of meaning: Must we 'go Gricean'?" *Mind & Language*, 28(3):342–375.

Barsalou, L. W., 1983. "Ad hoc categories." *Memory & Cognition*, 11(3):211–227.

———, 1987. "The instability of graded structure: Implications for the nature of concepts."

- U. Neisser (ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*, CUP, pp. 101–140.
- , 1999. “Perceptual symbol systems.” *Behavioral and Brain Sciences*, 22(4):577–660.
- Bealer, G., 1993. “A solution to Frege’s puzzle.” *Philosophical Perspectives*, 7:17–60.
- Bennett, J., 1990. *Linguistic behaviour*. Hackett Publishing Company.
- Bertolet, R., 1987. “Speaker reference.” *Philosophical Studies*, 52(2):199–226.
- Bezuidenhout, A., 2002. “Truth-conditional pragmatics.” *Philosophical Perspectives*, 16:105–134.
- Borg, E., 2004. *Minimal semantics*. OUP.
- , 2012. *Pursuing meaning*. OUP.
- Bortolotti, L., 2010. *Delusions and other irrational beliefs*. OUP.
- Brand, M., 1984. *Intending and acting: Toward a naturalized action theory*. MIT Press.
- Bratman, M., 1984. “Two faces of intention.” *The Philosophical Review*, 93(3):375–405.
- , 1987. *Intention, plans, and practical reason*. CUP.
- Brewer, B., 2011. *Perception and its objects*. OUP.
- Buchanan, R., 2010. “A puzzle about meaning and communication.” *Noûs*, 44(2):340–371.
- Burge, T., 1979. “Individualism and the mental.” *Midwest Studies in Philosophy*, 4:73–121.
- , 2010. *Origins of objectivity*. OUP.
- Burgess, A. & Sherman, B., 2014. “Introduction: A plea for the metaphysics of meaning.” A. Burgess & B. Sherman (eds.), *Metasemantics: New essays on the foundations of meaning*, OUP, pp. 1–16.
- Byrne, A., 2005. “Perception and conceptual content.” E. Sosa & M. Steup (eds.), *Contemporary debates in epistemology*, Blackwell, pp. 231–250.
- Camp, E., 2006. “Contextualism, metaphor and what is said.” *Mind & Language*, 21(3):280–309.
- , 2012. “Sarcasm, pretense, and the semantics/pragmatics distinction.” *Noûs*, 46(4):587–634.
- Camp, J. L., 2002. *Confusion: A study in the theory of knowledge*. HUP.
- Campbell, J., 1987. “Is sense transparent?” *Aristotelian Society Supplementary Volume*, 88:273–292.
- , 2002. *Reference and consciousness*. Clarendon.
- Cappelen, H., 2012. *Philosophy without intuitions*. OUP.
- Cappelen, H. & Lepore, E., 2005. *Insensitive semantics*. Blackwell.
- Carroll, D., 2007. *Psychology of language*. Cengage Learning, 5 edn.
- Carruthers, P., 1996. *Language, thought and consciousness*. CUP.

- Carston, R., 1988. "Implicature, explicature, and truth-theoretic semantics." R. Kempson (ed.), *Mental representations*, CUP, pp. 155–181.
- , 1998. "The semantics/pragmatics distinction: a view from relevance theory." *UCL Working Papers in Linguistics*, 10:1–30.
- , 2002. *Thoughts and utterances*. Blackwell.
- Chomsky, N., 1977. *Essays on form and interpretation*. North-Holland.
- , 1980. *Rules and representations*. CUP.
- Churchland, P. M., 1979. *Scientific realism and the plasticity of mind*. CUP.
- Ciecierski, T., 2009. "The multiple proposition approach reconsidered." *Logique et Analyse*, 52(208):423–440.
- Crane, T., 2009. "Is perception a propositional attitude?" *The Philosophical Quarterly*, 59(236):452–469.
- Davidson, D., 1984. *Inquiries into truth and interpretation*. OUP.
- , 1985. "Replies to essays I-IX." B. Vermazen & M. B. Hintikka (eds.), *Essays on Davidson: Actions and events*, Clarendon, pp. 195–229.
- , 1986. "A nice derangement of epitaphs." R. Grandy & R. Warner (eds.), *Philosophical grounds of rationality: Intentions, categories, ends*, Clarendon Press, pp. 157–174. Repr. in Davidson (2005), pp. 89–107.
- , 2005. *Truth, language, and history*. Clarendon Press.
- Dennett, D., 1978. *Brainstorms: Philosophical essays on mind and psychology*. MIT Press.
- , 1982. "Beyond belief." A. Woodfield (ed.), *Thought and object*, Clarendon, pp. 1–95. Repr. in Dennett (1987), pp. 117–202.
- , 1987. *The intentional stance*. MIT Press.
- , 1991. "Real patterns." *The Journal of Philosophy*, 88(1):27–51.
- Devitt, M., 1974. "Singular terms." *The Journal of Philosophy*, 71(7):183–205.
- , 1981a. *Designation*. CUP.
- , 1981b. "Donnellan's distinction." *Midwest Studies in Philosophy*, 6(1):511–526.
- , 1984. *Realism and truth*. Princeton University Press.
- , 1985. "Critical notice of Gareth Evans, *The varieties of reference*." *Australasian Journal of Philosophy*, 62(2):216–232.
- , 1989. "Against direct reference." *Midwest Studies in Philosophy*, 13:206–240.
- , 1991. *Realism and truth*. Princeton University Press, 2 edn.
- , 1996. *Coming to our senses: A naturalistic program for semantic localism*. CUP.
- , 2004. "The case for referential descriptions." M. Reimer & A. Bezuidenhout (eds.),

- Descriptions and beyond*, OUP, pp. 280–306.
- , 2006. *Ignorance of language*. OUP.
- , 2010. *Putting metaphysics first: Essays on metaphysics and epistemology*. OUP.
- , 2012a. “The role of intuitions.” G. Russell & D. G. Fara (eds.), *The Routledge companion to the philosophy of language*, Routledge, pp. 554–565.
- , 2012b. “Still against direct reference.” R. Schantz (ed.), *Prospects for meaning*, de Gruyter, pp. 61–84.
- , 2013a. “Is there a place for truth-conditional pragmatics?” *Teorema: Revista internacional de filosofia*, 32(2):85–102.
- , 2013b. “The myth of the problematic *de se*.” N. Feit & A. Capone (eds.), *Attitudes de se: Linguistics, epistemology, metaphysics*, CSLI Publications, pp. 133–162.
- , 2013c. “Three methodological flaws of linguistic pragmatism.” C. Penco & F. Domaneschi (eds.), *What is said and what is not: The semantics/pragmatics interface*, CSLI Publications, pp. 285–300.
- , 2013d. “What makes a property “semantic”?” A. Capone, F. L. Piparo, & M. Carapezza (eds.), *Perspectives on pragmatics and philosophy*, Springer, pp. 87–112.
- Devitt, M. & Sterelny, K., 1999. *Language and reality: An introduction to the philosophy of language*. MIT Press, 2 edn.
- Doerge, F. C., 2013. “Relevance theory and prolixity.” L. Goldstein (ed.), *Brevity*, OUP, pp. 299–312.
- Donnellan, K., 1966. “Reference and definite descriptions.” *The Philosophical Review*, 75(3):281–304. Repr. in Donnellan (2012), pp. 3–30.
- , 1968. “Putting Humpty Dumpty together again.” *The Philosophical Review*, 77(2):203–215. Repr. in Donnellan (2012), pp. 31–48.
- , 1970. “Proper names and identifying descriptions.” *Synthese*, 21(3-4):335–358. Repr. in Donnellan (2012), pp. 49–79.
- , 2012. *Essays on reference, language, and mind*. OUP.
- Downing, P., 1977. “On the creation and use of English compound nouns.” *Language*, 53(4):810–842.
- Dresner, E., 2006. “A measurement theoretic account of propositions.” *Synthese*, 153(1):1–22.
- Duchaine, B., 2011. “Developmental prosopagnosia: Cognitive, neural, and developmental investigations.” M. J. . J. H. Andy Calder, Gillian Rhodes (ed.), *The Oxford handbook of face perception*, OUP, pp. 821–838.
- Elbourne, P. D., 2005. *Situations and individuals*. MIT Press.
- Erard, M., 2008. *Um...: Slips, stumbles, and verbal blunders, and what they mean*. Random House LLC.

- Evans, G., 1979. "Reference and contingency." *The Monist*, 62(2):161–189. Repr. in Evans (1985), pp. 178–213.
- , 1981. "Understanding demonstratives." H. Parret & J. Bouveresse (eds.), *Meaning and understanding*, de Gruyter, pp. 280–303. Repr. in Evans (1985), pp. 291–321.
- , 1982. *The varieties of reference*. Clarendon.
- , 1985. *Collected papers*. Clarendon.
- Fara, D. G., 2015. "Names are predicates." *The Philosophical Review*, 124(1):59–117.
- Field, H., 1973. "Theory change and the indeterminacy of reference." *The Journal of Philosophy*, 70(14):462–481.
- Fiengo, R. & May, R., 1994. *Indices and identity*. MIT Press.
- , 2002. "Identity statements." G. Preyer & G. Peter (eds.), *Logical form and language*, OUP, pp. 169–203.
- , 2006. *De lingua belief*. MIT Press.
- Fine, K., 2007. *Semantic relationism*. Wiley-Blackwell.
- , 2010a. "Reply to Lawlor's 'Varieties of coreference'." *Philosophy and Phenomenological Research*, 81(2):496–501.
- , 2010b. "Semantic necessity." B. Hale & A. Hoffmann (eds.), *Modality: Metaphysics, logic, and epistemology*, OUP, pp. 65–80.
- Fodor, J., 1975. *The language of thought*. HUP.
- , 1978. "Propositional attitudes." *The Monist*, 61(4):501–523. Repr. in Fodor (1981), pp. 177–203.
- , 1981. *Representations: Philosophical essays on the foundations of cognitive science*. The Harvester Press.
- , 1985. "Fodor's guide to mental representation: The intelligent auntie's vade-mecum." *Mind*, 94(373):76–100. Repr. in Fodor (1990), pp. 3–30.
- , 1987. *Psychosemantics*. MIT Press.
- , 1990. *A theory of content and other essays*. MIT Press.
- , 2008. *LOT 2: The language of thought revisited*. OUP.
- Fodor, J. A., 1998. *Concepts: Where cognitive science went wrong*. Clarendon Press.
- Fromkin, V. A. (ed.), 1973. *Speech errors as linguistic evidence*. Mouton.
- García-Carpintero, M., 2001. "Gricean rational reconstructions and the semantics/pragmatics distinction." *Synthese*, 128(1-2):93–131.
- Gauker, C., 1997. "Domain of discourse." *Mind*, 106(421):1–32.
- , 2003. *Words without meaning*. MIT Press.

- , 2012. "Perception without propositions." *Philosophical Perspectives*, 26(1):19–50.
- Gibbs Jr, R. W. & Colston, H. L., 2012. *Interpreting figurative meaning*. CUP.
- Godfrey-Smith, P., 1994. "A continuum of semantic optimism." S. Stich & T. Warfield (eds.), *Mental representation*, Wiley-Blackwell, pp. 259–277.
- , 2004. "On folk psychology and mental representation." H. Clapin, P. Staines, & P. Slezak (eds.), *Representation in mind: New approaches to mental representation*, Elsevier, pp. 147–162.
- , 2005. "Folk psychology as a model." *Philosophers' Imprint*, 5(6):1–16.
- Grandy, R. & Warner, R., 1986. "Paul Grice: A view of his work." R. Grandy & R. Warner (eds.), *Philosophical grounds of rationality: Intentions, categories, ends*, Clarendon Press.
- Green, K., 2001. "Davidson's derangement: Of the conceptual priority of language." *Dialectica*, 55(3):239–258.
- Green, M. S., 2007. *Self-expression*. OUP.
- Greenberg, M., 2007. "Incomplete understanding, deference, and the content of thought." *UCLA School of Law Working Paper Series, Public Law & Legal Theory Working Paper 07-30*. <<http://ssrn.com/abstract=1030144>>.
- , 2014. "Troubles for content I." A. Burgess & B. Sherman (eds.), *Metasemantics: New essays on the foundations of meaning*, OUP, pp. 147–168.
- Grice, P., 1957. "Meaning." *The Philosophical Review*, 66(3):377–388. Repr. in Grice (1989), pp. 213–223.
- , 1969a. "Utterer's meaning and intention." *The Philosophical Review*, 78(2):147–177. Repr. in Grice (1989), pp. 86–116.
- , 1969b. "Vacuous names." D. Davidsson & J. Hintikka (eds.), *Words and objections: Essays on the work of W.V. Quine*, Reidel, pp. 118–145.
- , 1971. "Intention and uncertainty." *Proceedings of the British Academy*, 57:263–279.
- , 1975. "Logic and conversation." *Syntax and Semantics*, 3:41–58. Repr. in Grice (1989), pp. 22–40.
- , 1978. "Further notes on logic and conversation." *Syntax and Semantics*, 9:113–128. Repr. in Grice (1989), pp. 41–57.
- , 1982. "Meaning revisited." N. Smith (ed.), *Mutual knowledge*, Academic Press, pp. 223–243. Repr. in Grice (1989), pp. 283–303.
- , 1989. *Studies in the way of words*. HUP.
- Hacker, P. M. S., 2007. *Human nature: The categorial framework*. John Wiley & Sons.
- Haegeman, L., 1994. *Introduction to government and binding theory*. Blackwell.
- Hanfling, O., 2000. *Philosophy and ordinary language: The bent and genius of our tongue*. Routledge.
- Hanks, P. W., 2007. "The content–force distinction." *Philosophical Studies*, 134(2):141–164.

- , 2011. “Structured propositions as types.” *Mind*, 120(477):11–52.
- Harman, G., 1973. *Thought*. Princeton University Press.
- , 1976. “Practical reasoning.” *The Review of Metaphysics*, 29(3):431–463.
- Harris, D., 2014. *Speech-act-theoretic semantics*. Ph.D. thesis, CUNY Graduate Center.
- Hawthorne, J. & Manley, D., 2012. *The reference book*. OUP.
- Heck, R., 2002. “Do demonstratives have senses?” *Philosophers’ Imprint*, 2(2):1–33.
- Hestvik, A., 1995. “Reflexives and ellipsis.” *Natural Language Semantics*, 3(2):211–237.
- Jucker, A., 1994. “Irrelevant repetitions: A challenge to relevance theory.” *SPELL: Swiss Papers in English Language and Literature 7. Repetitions*, Gunter Narr Verlag, pp. 47–60.
- Kahneman, D., 2011. *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kaplan, D., 1978. “Dthat.” *Syntax and Semantics*, 9:383–400.
- , 1989a. “Afterthoughts.” J. Almog, J. Perry, & H. Wettstein (eds.), *Themes from Kaplan*, OUP, pp. 565–614.
- , 1989b. “Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals.” J. Almog, J. Perry, & H. K. Wettstein (eds.), *Themes from Kaplan*, OUP, pp. 481–563.
- , 1990. “Words.” *Proceedings of the Aristotelian Society*, (Supp. vol.) 64:95–119.
- Katz, J., 1972. *Semantic theory*. Harper and Row.
- King, J. C., 2013. “Supplementives, the coordination account, and conflicting intentions.” *Philosophical Perspectives*, 27(1):288–311.
- , 2014. “Speaker intentions in context.” *Noûs*, 48(2):219–237.
- Korta, K. & Perry, J., 2011. *Critical pragmatics: An inquiry into reference and communication*. CUP.
- Kripke, S., 1977. “Speaker’s reference and semantic reference.” *Midwest Studies in Philosophy*, 2:6–27. Repr. in Kripke (2011), pp. 99–124.
- , 1979. “A puzzle about belief.” A. Margalit (ed.), *Meaning and use*, Reidel, pp. 139–183. Repr. in Kripke (2011), pp. 125–161.
- , 1980. *Naming and necessity*. HUP.
- , 2011. *Philosophical troubles: Collected papers, vol. 1*. OUP.
- , 2013. *Reference and existence: The John Locke lectures*. OUP.
- Lance, M. N. & O’Leary-Hawthorne, J., 1997. *The grammar of meaning: Normativity and semantic discourse*. CUP.
- Lawlor, K., 2001. *New thoughts about old things: Cognitive policies as the ground of singular concepts*. Taylor & Francis.

- , 2005. "Confused thought and modes of presentation." *The Philosophical Quarterly*, 55(218):21–36.
- , 2007. "A notional worlds approach to confusion." *Mind & Language*, 22(2):150–172.
- , 2010. "Varieties of coreference." *Philosophy and Phenomenological Research*, 81(2):485–495.
- Lepore, E. & Stone, M., 2015. *Imagination and convention: Distinguishing grammar and inference in language*. OUP.
- Levelt, W. J. M., 1993. *Speaking: From intention to articulation*. MIT Press.
- Lewis, D., 1980. "Index, context, and content." S. Kanger & S. Öhman (eds.), *Philosophy and grammar*, Reidel, pp. 79–100. Repr. in Lewis (1998), pp. 21–44.
- , 1998. *Papers in philosophical logic*. CUP.
- Loar, B., 1976. "The semantics of singular terms." *Philosophical Studies*, 30(6):353–377.
- , 1986. *Mind and meaning*. CUP.
- Lockwood, M., 1975. "On predicating proper names." *The Philosophical Review*, 84(4):471–498.
- Ludlow, P., 2014. *Living words: Meaning underdetermination and the dynamic lexicon*. OUP.
- MacKay, A. F., 1968. "Mr. Donnellan and Humpty Dumpty on Referring." *The Philosophical Review*, 77(2):197–202.
- Mallon, R., Machery, E., Nichols, S., & Stich, S., 2009. "Against arguments from reference." *Philosophy and Phenomenological Research*, 79(2):332–356.
- Marcus, R. B., 1961. "Modalities and intensional languages." *Synthese*, 13(4):303–322. Repr. in Marcus (1993), pp. 3–23.
- , 1981. "A proposed solution to a puzzle about belief." *Midwest Studies in Philosophy*, 6:501–510.
- , 1983. "Rationality and believing the impossible." *The Journal of Philosophy*, 80(6):321–338. Repr. in Marcus (1993), pp. 143–161.
- , 1990. "Some revisionary proposals about belief and believing." *Philosophy and Phenomenological Research*, 50 (Supplement):133–153. Repr. in Marcus (1993), pp. 233–255.
- , 1993. *Modalities: Philosophical essays*. OUP.
- Matthews, R. J., 2007. *The measure of mind: Propositional attitudes and their attribution*. OUP.
- McDowell, J. H., 1980. "Meaning, communication, and knowledge." *Philosophical subjects: Essays presented to P. F. Strawson*, Clarendon Press, pp. 117–39. Repr. in McDowell (1998), pp. 29–50.
- , 1994/1996. *Mind and world*. HUP, 2 edn.
- , 1998. *Meaning, knowledge, and reality*. HUP.
- McGinn, C., 1981. "The mechanism of reference." *Synthese*, 49(2):157–186.

- Mele, A., 1992. *Springs of action: Understanding intentional behavior*. OUP.
- , 2001. “Acting intentionally: Probing folk notions.” B. Malle, L. Moses, & D. Baldwin (eds.), *Intentions and intentionality: Foundations of social cognition*, MIT Press, pp. 27–43.
- Michaelson, E., 2013. *This and that: A theory of reference for names, demonstratives, and things in between*. Ph.D. thesis, UCLA.
- Millikan, R. G., 1984. *Language, thought, and other biological categories: New foundations for realism*. MIT Press.
- , 1989a. “Biosemantics.” *The Journal of Philosophy*, pp. 281–297. Repr. in Millikan (1993), pp. 83–102.
- , 1989b. “In defense of proper functions.” *Philosophy of Science*, 56(2):288–302. Repr. in Millikan (1993), pp. 13–30.
- , 1993. *White queen psychology and other essays for Alice*. MIT Press.
- , 1994. “On unclear and indistinct ideas.” *Philosophical Perspectives*, 8:75–100.
- , 1997. “Images of identity: In search of modes of presentation.” *Mind*, 106(423):499–519.
- , 2000. *On clear and confused ideas: An essay about substance concepts*. CUP.
- Moltmann, F., 2003. “Propositional attitudes without propositions.” *Synthese*, 135(1):77–118.
- Murday, B., 2014. “Definite descriptions and semantic pluralism.” *Philosophical Papers*, 42(2):255–284.
- Neale, S., 1990. *Descriptions*. MIT Press.
- , 1992. “Paul Grice and the philosophy of language.” *Linguistics and Philosophy*, 15(5):509–559.
- , 1999. “Coloring and composition.” K. Murasugi & R. Stainton (eds.), *Philosophy and linguistics*, Westview, pp. 35–82.
- , 2001. “Implicature and coloring.” G. Cosenza (ed.), *Paul Grice’s heritage*, Brepols, pp. 135–180.
- , 2004. “This, that, and the other.” M. Reimer & A. Bezuidenhout (eds.), *Descriptions and beyond*, Clarendon, pp. 68–182.
- , 2005. “Pragmatism and binding.” Z. G. Szabó (ed.), *Semantics versus pragmatics*, Clarendon, pp. 165–285.
- , 2007. “On location.” M. O’Rourke & C. Washington (eds.), *Situating semantics: Essays on the philosophy of John Perry*, MIT Press, pp. 251–393.
- , 2008. “Term limits revisited.” *Philosophical Perspectives*, 22:375–442.
- , forthcoming. “Implicit reference.” G. Ostertag (ed.), *Meanings and other things*, MIT Press.
- Origg, G. & Sperber, D., 2000. “Evolution, communication and the proper function of language.” P. Carruthers & A. Chamberlain (eds.), *Evolution and the human mind*, CUP, pp. 140–169.

- Ostertag, G., 1998. "Introduction." G. Ostertag (ed.), *Definite descriptions*, MIT Press, pp. 1–33.
- , 1999. "A scorekeeping error." *Philosophical Studies*, 96(2):123–146.
- , 2013. "The 'Gödel' effect." *Philosophical Studies*, 166(1):65–82.
- Papineau, D., 1984. "Representation and explanation." *Philosophy of Science*, 51(4):550–572.
- , 1987. *Reality and representation*. Basil Blackwell.
- Partee, B. & Bach, E., 1984. "Quantification, pronouns, and VP anaphora." J. Groenendijk, T. Janssen, & M. Stokhof (eds.), *Truth, interpretation and information*, Dordrecht, pp. 99–130.
- Peacocke, C., 1983. *Sense and content: Experience, thought, and their relations*. Clarendon.
- Pears, D., 1985. "Intention and belief." B. Vermazen & M. B. Hintikka (eds.), *Essays on Davidson: Actions and events*, Clarendon, pp. 75–88.
- Perry, J., 1979. "The problem of the essential indexical." *Noûs*, 13:3–21. Repr. in Perry (2000), pp. 27–44.
- , 1986. "Thought without representation." *Proceedings of the Aristotelian Society*, (Supp. vol.) 60:137–152. Repr. in Perry (2000), pp. 171–188.
- , 1997. "Rip Van Winkle and other characters." *European Review of Philosophy*, 2:13–40. Repr. in Perry (2000), pp. 355–376.
- , 2000. *The problem of the essential indexical and other essays: Expanded edition*. CSLI Publications.
- , 2009. "Directing intentions." J. Almog & P. Leonardi (eds.), *The philosophy of David Kaplan*, OUP, pp. 187–202.
- , 2012. *Reference and reflexivity*. CSLI Publications. 2nd ed.
- Pietroski, P., 1992. "Intentionality and teleological error." *Pacific Philosophical Quarterly*, 73(3):267–282.
- , 2005. *Events and semantic architecture*. OUP.
- , 2007. "Systematicity via monadicity." *Croatian Journal of Philosophy*, 7(21):343–374.
- , 2011. "Minimal semantic instructions." C. Boeckx (ed.), *The Oxford handbook of linguistic minimalism*, OUP, pp. 472–498.
- Predelli, S., 2005. *Contexts: Meaning, truth, and the use of language*. Clarendon.
- Priest, G., 2006. *Doubt truth to be a liar*. Clarendon.
- Prinz, J. J., 2002. *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.
- Putnam, H., 1973. "Meaning and reference." *The Journal of Philosophy*, 70(19):699–711.
- Quine, W. V. O., 1960. *Word and object*. MIT Press.
- Recanati, F., 1986. "On defining communicative intentions." *Mind & Language*, 1(3):213–241.
- , 1989. "The pragmatics of what is said." *Mind & Language*, 4(4):295–329.

- , 1993. *Direct reference*. Blackwell.
- , 2001. “What is said.” *Synthese*, 128(1):75–91.
- , 2002. “Unarticulated constituents.” *Linguistics and Philosophy*, 25(3):299–345.
- , 2004. *Literal meaning*. CUP.
- , 2010. *Truth-conditional pragmatics*. OUP.
- , 2012. *Mental files*. OUP.
- Reimer, M., 1991a. “Demonstratives, demonstrations, and demonstrata.” *Philosophical Studies*, 63(2):187–202.
- , 1991b. “Do demonstrations have semantic significance?” *Analysis*, 51(4):177–183.
- , 1992a. “Demonstrating with descriptions.” *Philosophy and Phenomenological Research*, 52(4):877–893.
- , 1992b. “Three views of demonstrative reference.” *Synthese*, 93(3):373–402.
- , 1998. “Donnellan’s distinction/Kripke’s test.” *Analysis*, 58(2):89–100.
- , 2002. “Do adjectives conform to compositionality?” *Philosophical Perspectives*, 36(16):183–198.
- , 2004. “What malapropisms mean: A reply to Donald Davidson.” *Erkenntnis*, 60(3):317–334.
- Richard, M., 1983. “Direct reference and ascriptions of belief.” *Journal of Philosophical Logic*, 12(4):425–452.
- , 2013. “Marcus on belief and belief in the impossible.” *Theoria*, 28(3):407–420.
- Rosenthal, D. M., 1986. “Intentionality.” *Midwest Studies in Philosophy*, 10:151–184. Repr. in Rosenthal (2005), pp. 71–99.
- , 1989. “Postscript to ‘Intentionality’.” *Rerepresentations: Readings in the philosophy of mental representation*, Kluwer Academic Publishers, pp. 341–344. Repr. in Rosenthal (2005), pp. 100–102.
- , 2005. *Consciousness and mind*. OUP.
- Rothschild, D. & Segal, G., 2009. “Indexical predicates.” *Mind & Language*, 24(4):467–493.
- Russell, B., 1905. “On denoting.” *Mind*, 14(56):479–493.
- Sacks, O., 2010. *The mind’s eye*. Vintage.
- Sainsbury, M., 2002. *Departing from Frege: Essays in the philosophy of language*. Routledge.
- Salmon, N., 1986a. *Frege’s puzzle*. Ridgeview Publishing Company.
- , 1986b. “Reflexivity.” *Notre Dame Journal of Formal Logic*, 27(3):401–430. Repr. in Salmon (2007), pp. 32–57.
- , 1989. “How to become a Millian heir.” *Noûs*, 23(2):211–220.

- , 1991. “The pragmatic fallacy.” *Philosophical Studies*, 63(1):83–97. Repr. in Salmon (2007), pp. 298–308.
- , 1992. “Reflections on reflexivity.” *Linguistics and Philosophy*, 15(1):53–63. Repr. in Salmon (2007), pp. 58–66.
- , 2007. *Content, cognition, and communication: Philosophical papers II*. Clarendon.
- Saul, J. M., 1997. “Substitution and simple sentences.” *Analysis*, 57(2):102–108.
- , 1998. “The pragmatics of attitude ascription.” *Philosophical Studies*, 92(3):363–389.
- , 2002. “Speaker meaning, what is said, and what is implicated.” *Noûs*, 36(2):228–248.
- , 2007. *Simple sentences, substitution, and intuitions*. OUP.
- Schiffer, S., 1972. *Meaning*. OUP.
- , 1977. “Naming and knowing.” *Midwest Studies in Philosophy*, 2:28–41.
- , 1978. “The basis of reference.” *Erkenntnis*, 13(1):171–206.
- , 1981. “Indexicals and the theory of reference.” *Synthese*, 49(1):43–100.
- , 1987. *Remnants of meaning*. CUP.
- , 1988. “Review of Gareth Evans, *The varieties of reference*.” *The Journal of Philosophy*, 85(1):33–42.
- , 1992. “Belief ascription.” *The Journal of Philosophy*, 89(10):499–521.
- , 2003. *The things we mean*. Clarendon.
- , 2014. “Meaning and formal semantics in generative grammar.” *Erkenntnis*, 80(1):61–87.
- Schwarz, D. S., 1976. “Referring, singular terms, and presupposition.” *Philosophical Studies*, 30(1):63–74.
- Schwitzgebel, E., 2013. “A dispositional approach to the attitudes: thinking outside of the belief box.” N. Nottelmann (ed.), *New essays on belief: Constitution, content and structure*, Palgrave, pp. 75–99.
- Scott-Phillips, T., 2015. *Speaking our minds: Why human communication is different, and how language evolved to make it special*. Palgrave Macmillan.
- Searle, J., 1969. *Speech acts: An essay in the philosophy of language*. CUP.
- , 1975. “Indirect speech acts.” *Syntax and Semantics*, 3:59–82.
- , 1978. “Literal meaning.” *Erkenntnis*, 13(1):207–224.
- , 1979a. *Expression and meaning*. CUP.
- , 1979b. “Referential and attributive.” *The Monist*, 62(2):190–208. Repr. in Searle (1979a), pp. 137–161.
- , 1980. “The background of meaning.” J. Searle, F. Kiefer, & M. Bierwisch (eds.), *Speech act theory and pragmatics*, Reidel, pp. 221–232.

- , 1983. *Intentionality: An essay in the philosophy of mind*. CUP.
- Sennet, A., 2011. "Ambiguity." E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Summer 2011 edn.
- Shoemaker, S., 1995. "Moore's paradox and self-knowledge." *Philosophical Studies*, 77(2):211–228.
- Siegel, S., 2002. "The role of perception in demonstrative reference." *Philosophers' Imprint*, 2(1):1–21.
- , 2012. *The contents of visual experience*. OUP.
- Soames, S., 1989a. "Direct reference and propositional attitudes." J. Almog, J. Perry, & H. K. Wettstein (eds.), *Themes from Kaplan*, OUP, pp. 393–419.
- , 1989b. "Pronouns and propositional attitudes." *Proceedings of the Aristotelian Society*, 90:191–212.
- , 1989c. "Review of Gareth Evans, *Collected papers*." *The Journal of Philosophy*, 86(3):141–156.
- , 1994. "Attitudes and anaphora." *Philosophical Perspectives*, 8:251–272. Repr. in Soames (2009c), pp. 111–135.
- , 2002. *Beyond rigidity: The unfinished semantic agenda of naming and necessity*. OUP.
- , 2005. "Naming and asserting." Z. G. Szabó (ed.), *Semantics versus pragmatics*, OUP, pp. 356–382.
- , 2009a. "The gap between meaning and assertion: Why what we literally say often differs from what our words literally mean." *Philosophical essays, Vol. 1: Natural language: What it means and how we use it*, Princeton University Press, pp. 278–297.
- , 2009b. *Philosophical essays, Vol. 1: Natural language: What it means and how we use it*. Princeton University Press.
- , 2009c. *Philosophical essays, Vol. 2: The philosophical significance of language*. Princeton University Press.
- , 2010a. *Philosophy of language*. Princeton University Press.
- , 2010b. *What is meaning?* Princeton University Press.
- , 2014. "Cognitive propositions." J. C. King, S. Soames, & J. Speaks (eds.), *New thinking about propositions*, OUP, pp. 91–124.
- Speaks, J., forthcoming. "The role of speaker and hearer in the character of demonstratives." *Mind*.
- Sperber, D., 2000. "Metarepresentations in an evolutionary perspective." D. Sperber (ed.), *Metarepresentations: A multi-disciplinary perspective*, OUP, pp. 117–137.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G., & Wilson, D., 2010. "Epistemic vigilance." *Mind & Language*, 25(4):359–393.

- Sperber, D. & Wilson, D., 1986. *Relevance: Communication and cognition*. Blackwell.
- , 1995. *Relevance: Communication and cognition*. Blackwell, 2 edn.
- , 1998. “The mapping between the mental and the public lexicon.” P. Carruthers & J. Boucher (eds.), *Language and thought: Interdisciplinary themes*, CUP, pp. 184–200. Repr. in Wilson & Sperber (2012), pp. 31–46.
- , 2005. “Pragmatics.” F. Jackson & M. Smith (eds.), *The Oxford handbook of contemporary philosophy*, OUP, pp. 468–501. Repr. in Wilson & Sperber (2012), pp. 1–27.
- , unpub. ms. “Beyond speaker’s meaning.” Presented at the NYU Mind & Language Seminar, April 16, 2013.
- Stainton, R., forthcoming. “A deranged argument against public languages.” *Inquiry*.
- Stalnaker, R., 1984. *Inquiry*. MIT Press.
- Stanley, J., 2007. *Language in context*. Clarendon.
- Stine, G. C., 1978. “Meaning other than what we say and referring.” *Philosophical Studies*, 33(4):319–337.
- Strawson, P. F., 1950. “On referring.” *Mind*, 59(235):320–344. Repr. in Strawson (1971), pp. 1–27.
- , 1964. “Intention and convention in speech acts.” *The Philosophical Review*, 73(4):439–460. Repr. in Strawson (1971), pp. 149–169.
- , 1971. *Logico-linguistic papers*. Methuen.
- , 1974. *Subject and predicate in logic and grammar*. Ashgate Publishing, 2 edn. Pub. 2004.
- Thalberg, I., 1972. *Enigmas of agency: Studies in the philosophy of human action*. Allen and Unwin.
- Thau, M., 2002. *Consciousness and cognition*. OUP.
- Thráinsson, H., 2005. *Setningar. Handbók um íslenska setningafræði. Íslensk tunga, 3. bindi*. Almenna bókafélagið.
- , 2007. *The syntax of Icelandic*. CUP.
- Tomasello, M., 2008. *Origins of human communication*. MIT Press.
- Travis, C., 1975. *Saying and understanding: A generative theory of illocutions*. Blackwell.
- , 1985. “On what is strictly speaking true.” *Canadian Journal of Philosophy*, 15(2):187–229.
- , 1989. *The uses of sense*. OUP.
- , 1991. “Critical Notice: Annals of Analysis.” *Mind*, 100(2):237–264. Repr. in Travis (2008), pp. 65–93.
- , 1997. “Pragmatics.” B. Hale & C. Wright (eds.), *A companion to the philosophy of language*, Blackwell, pp. 87–107. Repr. in Travis (2008), pp. 19–64.

- , 2004. “The silence of the senses.” *Mind*, 113(449):57–94.
- , 2008. *Occasion-sensitivity: Selected essays*. OUP.
- Tye, M., 1995. *Ten problems of consciousness: A representational theory of the phenomenal mind*. MIT Press.
- Unnsteinsson, E. G., 2014. “Compositionality and sandbag semantics.” *Synthese*, 191(14):3329–3350.
- Velleman, J. D., 1989. *Practical reflection*. Princeton University Press.
- Wallace, D. F., 2006. “Authority and American usage.” *Consider the lobster and other essays*, Little, Brown and Company, pp. 66–127.
- Weiskopf, D. A., 2007. “Compound nominals, context, and compositionality.” *Synthese*, 156(1):161–204.
- Wettstein, H. K., 1979. “Indexical reference and propositional content.” *Philosophical Studies*, 36(1):91–100. Repr. in Wettstein (1991), pp. 20–28.
- , 1981. “Demonstrative reference and definite descriptions.” *Philosophical Studies*, 40(2):241–257. Repr. in Wettstein (1991), pp. 35–49.
- , 1984. “How to bridge the gap between meaning and reference.” *Synthese*, 58(1):63–84. Repr. in Wettstein (1991), pp. 69–85.
- , 1991. *Has semantics rested on a mistake?* Stanford University Press.
- Wharton, T., 2009. *Pragmatics and non-verbal communication*. CUP.
- Wiggins, D., 1975. “Identity, designation, essentialism and physicalism.” *Philosophia*, 5(1):1–30.
- Williams, E. S., 1977. “Discourse and logical form.” *Linguistic Inquiry*, 8(1):101–139.
- Williamson, T., 2000. *Knowledge and its limits*. OUP.
- Wilson, D., 2000. “Metarepresentation in linguistic communication.” D. Sperber (ed.), *Metarepresentations: A multi-disciplinary perspective*, OUP, pp. 411–448. Repr. in Wilson & Sperber (2012), pp. 230–258.
- , 2005. “New directions for research on pragmatics and modularity.” *Lingua*, 115(8):1129–1146.
- Wilson, D. & Sperber, D., 1993. “Linguistic form and relevance.” *Lingua*, 90(1):1–25. Repr. in Wilson & Sperber (2012), pp. 149–168.
- , 2002a. “Pragmatics, modularity, and mindreading.” *Mind & Language*, 17(1–2):3–23. Repr. in Wilson & Sperber (2012), pp. 261–278.
- , 2002b. “Truthfulness and relevance.” *Mind*, 111(443):583–632. Repr. in Wilson & Sperber (2012), pp. 47–83.
- , 2012. *Meaning and relevance*. CUP.
- Wittgenstein, L., 1953. *Philosophical investigations*. Wiley-Blackwell, 4 edn. Trans. by G. E. M. Anscombe, P. M. S. Hacker, J. Schulte (2007).