

**UCLA**

**UCLA Electronic Theses and Dissertations**

**Title**

This and That: A Theory of Reference for Names, Demonstratives, and Things in Between

**Permalink**

<https://escholarship.org/uc/item/6bw3472d>

**Author**

Michaelson, Eliot

**Publication Date**

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA  
Los Angeles

This and That: A Theory of Reference for Names, Demonstratives, and  
Things in Between

A dissertation submitted in partial satisfaction of the  
requirements for the degree Doctor of Philosophy  
in Philosophy

by

Eliot Michaelson

2013

© Copyright by  
Eliot Michaelson  
2013

## ABSTRACT OF THE DISSERTATION

This and That: A Theory of Reference for Names, Demonstratives, and  
Things in Between

by

Eliot Michaelson

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2013

Professor Sam Cumming, Co-Chair

Professor David Kaplan, Co-Chair

According to intentionalism, the meaning of my words is determined by my intentions in using them. This theory, while both appealingly simple and surprisingly accurate, quickly runs into a problem: according to intentionalism, it looks like my words mean whatever I want them to mean. But surely there are some limits! We should, for instance, hardly want to follow Lewis Carroll's Humpty Dumpty in saying that 'glory' can mean 'a nice knock-down argument' if only one really wants it to. And yet, when we strive to offer a semantics for terms like demonstratives (e.g. 'this' and 'that') and pronouns (e.g. 'he', 'she', and 'it'), appealing to speakers' intentions has struck many philosophers not just as an appealing option, but as the only viable option around.

The primary goal of this thesis is to introduce a satisfying, and genuinely non-Humpty Dumpty, version of intentionalism. The view for which I argue—what I call the 'constraint theory'—avoids two serious problems facing earlier attempts to overcome Humpty Dumpty-

ism. First, many earlier intentionalist theories make implausible predictions in cases where the speaker is sufficiently confused about the context. And, second, most earlier theories fail to generalize beyond demonstratives and pronouns. That is, these theories have been so tailored to account for the particular characteristics of these terms that they make inaccurate predictions when extended to account for other sorts of referential terms, like names. In contrast, the constraint theory not only makes accurate predictions in cases of speaker confusion, it looks well-situated to generalize beyond demonstratives and pronouns—so as to provide a unified account of the meanings of referential terms in context.

Chapter 1 introduces the problem of reference fixing in more detail, and discusses a bit of its history. While the Humpty Dumpty Problem was first introduced as an objection to the claim that there are distinctly referential uses of definite descriptions, I explain why the problem turns out to be even worse for intentionalist accounts of demonstratives. These terms will thus serve as the focus of the three main chapters.

Chapter 2 considers the prospects for defending a non-intentionalist theory of how the reference of demonstratives is fixed in context. By far the most popular theory of this sort is ‘salientism’, or the view that whatever is maximally salient in context is what a use of a demonstrative refers to. Maximal salience, in turn, can be spelled out in several different ways. I argue that each of these ways exhibits serious flaws, and that there are reasons to think that no version of salience will be able to account for certain sorts of cases—involving speakers who are trying to deceive their listeners in a particular sort of way.

Chapter 3 introduces a classic hard case for intentionalists, the ‘Carnap-Agnew’ case. The force of the case requires our imagining a speaker who is confused about the world in a particular sort of way—and it is this confusion, in particular, that makes the case a hard one for intentionalists. I consider three intentionalist theories that purport to account for this sort of case: the perceptual grounding theory, the coordination account, and the neo-Gricean account. I argue that, while each can account for at least some versions of the Carnap-Agnew case, each suffers from some serious drawbacks. The perceptual grounding

theory is extremely limited in its scope, the coordination account relies on several implausible assumptions about idealized listeners, and the neo-Gricean theory seems forced to collapse linguistic and nonlinguistic knowledge in order to make accurate predictions. I then introduce my own proposal, the constraint theory, and show how it can account not just for the Carnap-Agnew case and its many variants, but also for analogous cases involving names.

Chapter 4 focuses on the neo-Gricean theory of reference. One possible concern with the criticism of that theory offered in chapter 3 is that it might appear to rest on what amounts to a mere technical flaw. That flaw might, in turn, seem to be fixable. In this chapter, I therefore endeavor to introduce a deeper problem for the neo-Gricean—one that should prove instructive for the theory of meaning more generally. Contrary to what the neo-Gricean predicts, speakers can both intend to and succeed in using names and demonstratives to deceive their listeners not just about the world, but about the very content of what they say. I call these ‘sneaky’ uses of referential terms, and I spend most of the chapter working through the ramifications of these cases for the Gricean theory of meaning and its successors. I argue that these cases present a serious challenge to that class of theories, and one that is not easily avoided. The constraint theory, in contrast, neatly accounts for sneaky cases. What’s more, it does so while still preserving many of the virtues of the Gricean theory.

Chapter 5 concludes the thesis by briefly tying up two loose ends: showing how the theory can be extended to account for so-called ‘true’ indexicals like ‘I’, ‘here’, and ‘now’, and explaining how the theory can endorse the claim that sneaky uses of referential terms partially ground the standing meaning of these very terms while simultaneously endorsing the claim that sneaky uses of referential terms asymmetrically depend on cooperative uses.

The dissertation of Eliot Michaelson is approved.

Edward Keenan

Sam Cumming, Committee Co-Chair

David Kaplan, Committee Co-Chair

University of California, Los Angeles

2013

in memory of my uncle Len Kunin and my dog Cassidy

# Contents

- 1 Introduction** **1**
  - 1.1 The Basic Question . . . . . 1
  - 1.2 Humpty Dumpty and His Discontents . . . . . 3
  - 1.3 The Shape of the Arguments to Come . . . . . 7
  
- 2 Against Salientism** **13**
  - 2.1 Introduction . . . . . 13
  - 2.2 The Problem of Reference Fixing . . . . . 15
  - 2.3 Salientism . . . . . 16
  - 2.4 Arguments for Salientism . . . . . 20
    - 2.4.1 The Argument from Cases . . . . . 20
    - 2.4.2 The Argument from Recoverability . . . . . 22
  - 2.5 Why These Arguments Fail . . . . . 23
    - 2.5.1 Why the Argument from Cases Fails . . . . . 24
    - 2.5.2 Why the Argument from Recoverability Fails . . . . . 28
  - 2.6 Why Salientism Should be Rejected . . . . . 30
  - 2.7 Conclusion . . . . . 37
  
- 3 Reference and Confusion** **39**
  - 3.1 Introduction . . . . . 39
  - 3.2 Simple Intentionalism . . . . . 43

3.3	Two Cases of Speaker Confusion . . . . .	45
3.3.1	The Carnap-Agnew Case . . . . .	45
3.3.2	The Smith-Jones Case . . . . .	47
3.4	Sophisticated Intentionalisms . . . . .	49
3.4.1	The Occurrent Perception Theory . . . . .	49
3.4.2	The Coordination Account . . . . .	51
3.4.3	Structured Intentionalism . . . . .	55
3.5	Smith-Jones Revisited . . . . .	61
3.6	The Constraint Theory of Reference . . . . .	62
3.6.1	Communicative Plans and Signals . . . . .	63
3.6.2	Public Commitments and Filters . . . . .	65
3.6.3	A Guide to Handling Confusion and Absence . . . . .	70
3.7	Conclusion . . . . .	72
<b>4</b>	<b>Sneaky Reference</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	The (neo-)Gricean Theory of Meaning . . . . .	79
4.3	The neo-Gricean Theory of Reference . . . . .	82
4.4	Names and Referential Intentions . . . . .	85
4.5	Sneaky Intentions . . . . .	88
4.6	Defensive Maneuvers . . . . .	93
4.7	The Constraint Theory Revisited . . . . .	97
4.7.1	Cooperative Plans and Deceptive Plans . . . . .	97
4.7.2	Sneaky Reference and Lazy Reference . . . . .	100
4.8	Conclusion . . . . .	103
<b>5</b>	<b>Conclusion</b>	<b>107</b>
5.1	Summary . . . . .	107

5.2	Why I Still Can't be Napoleon, Despite Trying . . . . .	110
5.3	What Grounds Meaning? . . . . .	113
5.4	Closing Thoughts . . . . .	115
	<b>Bibliography</b>	<b>116</b>

# Acknowledgments

Thanks first and foremost to my co-chairs, David Kaplan and Sam Cumming. From David, I have learned how to do philosophy. I have also learned, repeatedly, that I am not nearly as clever as I think I am. I am deeply grateful to him for teaching me both these things. Sam has been my teacher, mentor, and friend through six of my seven years of graduate school. I would not have made it through without him. Nor would I be recognizable to myself as a philosopher. I could not have asked for a better pair of advisors.

I owe both an immense personal and professional debt to my third committee member from the UCLA philosophy department, Mark Greenberg. Mark has been stuck teaching me philosophy—at times, I believe, much to his dismay—since I was a first-year undergraduate at Deep Springs College. He has provided me excellent advice and some much needed encouragement over the years, in addition to an exceptional amount of scathing criticism. He has also introduced me to many of the subjects I love. For fifteen years, I have always known where I stand with Mark—and that is no small thing.

My two outside committee members also deserve to be singled out for thanks. Ed Keenan has had the particular misfortune of trying to teach me formal methods in linguistics, and the patience to do it well. Well enough, in fact, that there is basically no formalism in what follows. That, I take it, is a testament to Ed's determination in teaching me to separate issues which are best approached formally from those which are not. Jeff King, on the other hand, has taught me the value of careful, detailed philosophical theorizing. Time and time again, he has called out my bad arguments, and then helped me build up new ones—even when

they have been directed against him. I can only hope that he finds the current arguments I offer against his views slightly less unsatisfactory than the ones I started with.

In the course of writing this dissertation, I have incurred far too many philosophical debts to describe each in detail. Briefly, for many, many helpful conversations, comments on draft chapters, and more cribbed ideas than I would like to admit, thanks to: Zed Adams, Joseph Almog, Josh Armstrong, Matthew Babb, Kent Bach, Mel Bervoets, Sarah Bittman, Anthony Brueckner, Ray Buchannan, Tyler Burge, Alexi Burgess, Heather Burnett, Arudra Burra, Elizabeth Camp, Alejandro Pérez Carballo, Jonathan Cohen, Jorah Dannenberg, Shamik Dasgupta, Kenny Easwaran, Andy Egan, Ashley Feinsinger, David Friedell, Marina Folescu, Sarah Jansen, Nathaniel Hansen, Daniel Harris, Andrew Hsu, Rob Hughes, Robin Jeshion, Andrew Jewell, Rachel Johnson, Brent Kious, Krista Lawlor, Ben Lennertz, Karen Lewis, Hallie Liberto, Kathryn Lindeman, Dustin Locke, Yannig Luthra, Chauncey Maher, Lisa Miracchi, Matthew Moss, Sarah Murray, Stephen Neale, Paul Nichols, Tyke Nunez, Eileen Nutting, Jessica Pepp, David Plunkett, Gabe Rabin, Alexandru Radulescu, Peter Railton, Andrew Reisner, Jessica Rett, Karl Schafer, John Scott-Railton, Sam Shpall, Jeff Speaks, Will Starr, Luca Struble, Jesse Summers, Eric Swanson, Sabine Tsudura, Tom Ward, Tamar Weber, Steve White, and Jack Woods. For their generous support of this project in the form of a dissertation year fellowship, thanks is due as well to the Mellon Foundation and the American Council of Learned Societies.

I owe a longstanding philosophical debt to David Arndt, Jim Conant, Arnold Davidson, Michael Forster, John Haugeland, Kinch Hoekstra, Michael Kremer, and Josef Stern, who are collectively responsible for having gotten the philosophy bug into me as an undergraduate.

I am also deeply indebted to my friends for helping me survive grad school with some shred of a soul remaining. Particular thanks are due to Tom Garvey, for being there for me the since first grade, and to Tyke Nunez, for heading all the way down the philosophy rabbit-hole with me. The rest of you know who you are. Thank you.

Finally, thanks to my parents. I have never doubted your support for a minute.

# Vita

## Education

- M.A. Philosophy, University of California, Los Angeles, 2008
- B.A. Philosophy, University of Chicago, 2005
- A.A. Deep Springs College, 2001

## Publications

- “Shifty Characters,” *Philosophical Studies*, in press
- “Indexicality and the Answering Machine Paradox,” with Jonathan Cohen, *Philosophy Compass* 8(6): 580–592, 2013
- “Justice for Unicorns,” *Proceedings of the Aristotelian Society*, 112(3): 351–360, 2012

## Presentations

- “A Constraint Semantics for Demonstratives.” CUSP (California Universities Semantics and Pragmatics) IV. University of Southern California, Los Angeles, CA, February 2013
- “Knowing That not Knowing How.” UCLA Philosophy Department Albritton Society. University of California, Los Angeles, Los Angeles, CA, January 2013
- “Justice for Unicorns.” The Joint Session of Mind and The Aristotelian Society, Graduate Session. University of Sussex, Brighton, England, July 2012

- “The Lying Test.” Cornell University Semantics Group. Cornell University, Ithaca, NY, November 2012
- “Answering Machine Problems.” UCSD Semantics Babble. University of California, San Diego, San Diego, CA, May 2012
- “Shifty Characters.” New York Philosophy of Language Workshop. New York University, New York City, NY, April 2012
- “Lying and Content I.” UCLA Undergraduate Philosophy Club. University of California, Los Angeles, Los Angeles, CA, June 2011
- “Lying and Content I.” American Philosophical Association Pacific Meeting. San Diego, CA, April 2011
- “Lying and Content I.” Berkeley-Stanford-Davis Graduate Philosophy Conference. University of California, Davis, Davis, CA, April 2011

## **Fellowships, Honors, and Awards**

- Dissertation Completion Fellowship. Mellon Foundation/American Council of Learned Societies; 2012–2013
- Dean’s Humanities Fellowship. UCLA Graduate Division; 2011–2012
- Travel Grant. The Aristotelian Society; 2011
- Commendation for Teaching Excellence. UCLA Department of Philosophy; 2010
- Graduate Research Mentorship. UCLA Graduate Division; 2009–2010
- Graduate Summer Research Mentorship. UCLA Graduate Division; 2009
- Travel Grant. Association for Symbolic Logic; 2008
- Graduate Summer Research Mentorship. UCLA Graduate Division; 2007
- Dean’s Humanities Fellowship. UCLA Graduate Division; 2006–2007

# Chapter 1

## Introduction

*“The question is,” said Alice, “whether you can make words mean so many different things.”*  
*“The question is,” said Humpty Dumpty, “which is to be master—that’s all.”*  
*Lewis Carroll, Through the Looking Glass—1871*

### 1.1 The Basic Question

This dissertation aims to answer the question: what fixes the semantic values, or referents, of terms like proper names (hereafter, just ‘names’) and demonstratives (terms like ‘this’ and ‘that’) in context? Answering this question is crucial to the viability of the project of natural language semantics—at least as it is standardly understood. This is because, in the wake of Kaplan’s seminal work on indexicals and demonstratives (cf. Kaplan 1978a,b, 1989a,b), philosophers of language and linguistic semanticists alike have aspired to offer theories of meaning that explain not just what *sentence-types* mean, but also what *uses of sentences* mean in context.<sup>1</sup> This project thus involves assigning both standing, context-invariant meanings to terms, and also meanings *in context*. Those contextualized meanings can then be built up—according to semantic rules corresponding to the syntactic structure

---

<sup>1</sup>Note that this stands in explicit contrast to at least some prominent earlier work in philosophy of language—in particular, Montague (1974). See also Lewis (1975). For some earlier precedents to Kaplan’s work, see in particular Reichenbach (1947), Strawson (1950), and Donnellan (1966).

of the sentences in which they are embedded—into the meanings of sentences in context. These contextually-embedded sentence meanings are what allow us to predict the truth or falsity of sentences relative to contexts.<sup>2</sup> What’s more, it is not unreasonable to hope that these meanings will help us better understand, and perhaps even predict, how particular sentence-types can be used to communicate thoughts in context.

In response to this question of what fixes the semantic values of names and demonstratives in context, I argue for a novel form of intentionalism about meaning. ‘Intentionalism’ should here be understood as referring to a class of views unified by their common commitment to the following claim: it is speakers’ intentions (or other mental states) that fix meaning in context. Intentionalist theories have long faced the following sort of problem: since they bind meaning to speakers’ intentions, it can look as though speakers should be able to use words to mean whatever they want them to mean. But surely this cannot be correct! For example, we should not say, following Lewis Carroll’s Humpty Dumpty, that ‘glory’ can mean ‘a nice knock-down argument’ if only we want it to hard enough.<sup>3</sup> Rather, intentions must somehow fix meaning within the limits of reason. The problem is fleshing out what these ‘limits of reason’ amount to in enough detail to obtain a predictive theory—while also not making too many false predictions. Accordingly, most of the thesis is dedicated to exploring the variety of ways in which intentionalists might try to make good on this promissory note: both in terms of how they have to this point, and in terms of how I propose that they should.

In brief, I suggest that these limits of reason are imposed not by any general considerations regarding the nature of linguistic intentions, but rather by the conventions governing the use of particular referential terms. This material imposes certain bounds on how these terms can be used—not necessarily to direct the listener to a particular object, but rather to pick out a particular object in context. In some sense then, the thesis proposes that Searle (1969) was

---

<sup>2</sup>I will also sometimes speak of utterances, rather than sentences, as bearing meaning or truth-conditions in context, in instances where this fits better with the surrounding text. While it has been a matter of some debate whether these are equivalent and, if not, which is properly the bearer of meaning and truth, such concerns are largely irrelevant for our purposes here (cf. Kaplan 1989a,b, Bach 2005, and Stevens 2009).

<sup>3</sup>Cf. Lewis Carroll, *Through the Looking Glass*, chapter 6.

half right when he claimed that “The sentence...provides a conventional means of achieving the intention to produce a certain illocutionary effect in the hearer” (Searle 1969, p. 48). Sentences and their sub-parts do serve as conventional means of achieving certain intended effects. But not all of those intended effects have to do with listeners. In fact, speakers sometimes go to great lengths to prevent their listeners from recognizing the full range of effects that they are aiming to achieve with particular uses of terms and sentences.

## 1.2 Humpty Dumpty and His Discontents

Before turning to my main arguments in any detail, allow me to start at the beginning, with the sort of intentionalist theory first introduced in Donnellan (1966). Donnellan’s main concern in this paper was to argue for a distinction between ‘attributive’ and ‘referential’ uses of definite descriptions. Attributive uses are supposed to exhibit a logical form roughly along the lines of Russell (1905)’s suggestion: “the F is G” is true just in case there is one and only one F, and that F is also G. We, however, are more interested in Donnellan’s thoughts on referential uses. These, according to Donnellan, are instances of genuine, unmediated reference—with the reference of any particular use being determined, in context, by the speaker’s intentions, or, as Donnellan sometimes puts it, by her having some particular object ‘in mind’ (Donnellan 1966, pp. 285–86).<sup>4</sup>

In response to Donnellan’s proposal, MacKay (1968) points out that, unless something more is said, this view seems bound to devolve into ‘Humpty Dumptyism’ about referential uses of descriptions. That is, Donnellan’s view would seem to entail, implausibly, that speakers can use definite descriptions to refer to whatever they want, whenever they want. On MacKay’s understanding of Donnellan (1966), neither the descriptive material (‘F’) nor anything else constrains the sorts of intentions speakers are capable of having with regard

---

<sup>4</sup>For the sake of clarity, I will use the gendered pronoun ‘she’ to refer to speakers throughout this dissertation, and the gendered pronoun ‘he’ to refer to listeners—except when the specifics of the case under consideration dictate otherwise. I will use the possessive versions of each of these pronouns accordingly.

to the reference of definite descriptions—and hence what those descriptions are capable of referring to in context. A speaker might thus use the expression ‘the rock’ to refer to a book on the table. This, MacKay reasons, is patently absurd; Donnellan’s theory should thus be rejected (MacKay 1968, pp. 200-02).

In a response to MacKay, Donnellan substantially refines his earlier claims about referential uses of definite descriptions. Here is Donnellan (1968):

The fact about intentions that I want to stress is that they are essentially connected with expectations. Ask someone to flap his arms with the intention of flying. In response he can certainly wave his arms up and down, just as one can easily on command say the words “It’s cold here.” But this is not to do it with the intention of flying. Nor does it seem to me that a normal adult in normal circumstances can flap his arms and in doing so really have that intention. Perhaps one can, by a stretch of the imagination, conceive of someone (a child, say, who has seen birds flying) doing this. But such a person—the child, for example—would have expectations not shared with us. Similarly, one cannot say entirely out of the blue, “It’s cold here” and mean “It’s hot here,” but not, I think, because whatever one’s intentions the words will not get invested with that meaning. Rather, we can explain this by the impossibility of having the right intention in such circumstances. (Donnellan 1968, p. 212)

And again a bit later:

What MacKay has ignored is the fact that the intention to refer to something in using a definite description is a complex intention involving expectations regarding one’s audience. When a speaker uses a definite description referentially he intends his audience to take the description as characterizing what it is he wants to talk about. In so doing he hopes that they will successfully recognize what that is. (Donnellan 1968, p. 214)

So, according to Donnellan (1968): (i) speakers, like other agents, cannot intend to do what they expect not to be able to accomplish<sup>5</sup>; and (ii) intentions to refer are more accurately conceived of as intentions to elicit a certain reaction in one’s audience. Specifically, referential intentions are intentions for one’s audience to recognize some particular object as the referent. Interestingly, Donnellan’s response here essentially prefigures one of the ways of restricting

---

<sup>5</sup>Alternatively: one cannot intend to do what one takes it to be impossible to achieve.

intentionalism that we are going to focus on over the course of this thesis: the neo-Gricean theory.<sup>6</sup> According to that theory, a restriction very much like the one that Donnellan proposes is supposed to explain why, when speakers intend to use demonstratives in certain ways, they fail in securing a referent for those terms.

Returning briefly to Humpty Dumpty’s claim regarding the term ‘glory’, the analogous response would run as follows: although it is up to the speaker to determine what ‘glory’ means in context via her intentions, she cannot genuinely intend to use the term ‘glory’ to mean ‘a nice knock-down argument’—since she would know that the listener will be unable to recover this. A similar response is available with respect to MacKay (1968)’s case: a speaker cannot really intend to use ‘the rock’ to refer to a book. Again, the problem is that the speaker will not expect the listener to be able to recover her meaning here—at least without offering some further guidance with regard to her intentions. In both these cases, Donnellan claims, the speaker cannot even form an intention to use these terms in such deviant ways. Forming such an intention would require that one expect to have at least some chance of succeeding in what one is setting out to accomplish—which, according to Donnellan, amounts to making one’s meaning clear to one’s audience.

In the case of ‘glory’, it is easy to suppose that the problem here is one of violating the conventions associated with the use of that term. What is frustratingly lacking in Donnellan’s response to MacKay, however, is any sort of parallel explanation with regard to MacKay’s ‘the rock’ example. The problem is that Donnellan himself insists that ‘the F’ can in fact be used to refer to non-F’s. So what puts *this* use of the definite description beyond the bounds of reason, whereas one can use ‘the man drinking the martini’ to refer to a man drinking water from a martini glass even though he is not drinking a martini—at least according to Donnellan (1966, p. 287)? No answer is forthcoming. While I am not suggesting that we quibble with Donnellan’s claim that one cannot (usually, at least) use the expression ‘the

---

<sup>6</sup>Donnellan (1968) in fact acknowledges the similarity between his account and Grice (1989d)’s theory of meaning—from which the neo-Gricean theory is derived—at Donnellan (1968, p. 212).

rock' to refer to a book on the table, I think that we should hope to understand why this is the case, rather than just observing that it is.

But in fact the problem is worse than this—and in two distinct ways. First, while Donnellan and MacKay were interested primarily in definite descriptions, similar issues arise with regard to simple demonstratives like 'this' and 'that'. And whereas we might have hoped to be able to leverage the descriptive material of a definite description so as to constrain reference, simple demonstratives offer us little hope in this regard. Just consider how little a use of 'this' or 'that' indicates about the referent. One might therefore expect for these terms to be the sort of thing that can be used to refer to just about anything in just about any context. This is hardly the case, however; in fact, uses of 'this' and 'that' can only be used to refer to a fairly constrained set of things relative to a context. For example, one cannot point to a table that has nothing to do with Calvin Normore (i.e. one that does not belong to him, that he does not regularly sit on, etc.), utter the sentence "That has the longest beard in the UCLA philosophy department," and succeed in referring to Calvin Normore—even if one somehow manages to form the intention to refer to Normore with this use of 'that'.

Second, suppose one were to try to block this worry by claiming that it is simply impossible for a speaker to have such an intention. The explanation would seem to run: since the speaker knows that she cannot hope to prompt the listener to identify Normore as the referent with her use of 'that', she cannot intend to use that term to refer to Normore. But this is too quick. Notice that the speaker might have all sorts of false beliefs—none of which would impugn her competence as a speaker of English—that would serve to underwrite the reasonableness of this intention. She might, for instance, believe that Normore is a shape-changer with a proclivity for taking on the form of a table at this exact location. This would serve to underwrite the speaker's belief that her utterance will serve to help the listener identify the referent, at least from her own internal point of view. But even where the speaker succeeds in forming the intention to refer to Normore with her use of 'that', it should be perfectly clear that she does not succeed in referring to Normore in this instance.

### 1.3 The Shape of the Arguments to Come

At this point, I hope to have motivated the need, if we are to seriously consider adopting intentionalism, to say something about the bounds of reasonable use. Let me therefore step back for a minute to sketch the overall shape of the arguments to come.

Two methodological details are worth noting at the outset: first, I will focus primarily on demonstratives like ‘this’ and ‘that’. As far as I am aware, it is with regard to such terms that the best-developed versions of intentionalism, and the most detailed arguments for these particular views, are to be found. That said, chapters 3 and 4 will inquire into the semantics of names as well. The reason for this is that many of the best theories for terms like ‘this’ and ‘that’ clearly stumble when trying to account for the reference of names. My ‘constraint theory’, in contrast, accounts for uses of names equally well as it does for uses of demonstratives. That, I will argue, is a particular virtue of the theory. I do not mean to be assuming that the potential to generalize in this manner is essential to a viable theory of demonstrative reference; rather, what I am assuming is that *ceteris paribus* more general semantic theories are to be preferred to more narrow and specialized ones.

Second, while my focus here will be on isolating a the best version of intentionalism on offer, I will not begin by assuming that some version of intentionalism must be correct. Rather, I begin by offering several arguments against the two main alternatives to intentionalism: namely, ‘salientism’ and ‘demonstrationism’.<sup>7</sup> Salientist views are unified by their commitment to the claim that some type of salience—whatever exactly that amounts to—determines reference (cf. Wettstein 1984, Gauker 2008, Mount 2008). Demonstrationists, on the other hand, hold that it is the demonstrations which often accompany uses of demonstratives that somehow fix their reference (cf. Kaplan 1978a, McGinn 1981). Both sorts of theories face what I take to be fatal problems. That said, they are worth taking seriously,

---

<sup>7</sup>Somewhat surprisingly, other intentionalists have failed to offer serious arguments against the former position, and have argued in remarkably little detail against the latter. I therefore undertake a necessary, and long overdue, corrective here.

and investigating in detail, nonetheless—not least because they strike many as intuitive. What’s more, such theories offer a *prima facie* plausible explanation of a set of cases that have long proved problematic for intentionalists.

On the other hand, I will do little to address the relatively recent suggestion—which, although occasionally mooted, has rarely endorsed outright—that, really, there is no theory at all to be had about what fixes the reference of demonstratives in context (cf. Neale 2004, Heck 2013). The proper response to such speculation, it seems to me, is simply to offer a viable theory of demonstrative reference. Since the suggestion that there is no theory of demonstrative reference gains at least some of its appeal, I take it, from a pessimistic induction over previous, failed attempts to offer such a theory, if I can manage to offer a successful theory below, this should serve to undermine at least one of the motivations for this ‘no theory’ theory. The utility of the theory on offer would then serve to do at least a bit more work. To be clear, I make no claim to being able to settle this matter decisively here. Still, we must start somewhere.

With these preliminaries out of the way, let me briefly outline what is to come. I begin with the question of whether there is a viable alternative to intentionalism about demonstrative reference. I argue that both demonstrationism and salientism suffer from serious problems. Since these are the primary—and, indeed, the only well-developed—alternatives to intentionalism, I conclude that there is good reason to think no such alternative is available, at least at present. This makes it all the more pressing to sketch a viable version of intentionalism. I proceed to canvass the various extant proposals and to explain why each of them is ultimately unsatisfactory. Finally, I turn to the task of developing a better form of intentionalism—a theory that succeeds where its predecessors have failed. I will attempt to show that the constraint theory represents such an improved version of intentionalism.

In slightly more detail: in chapter 2, I begin by explaining why demonstrationism of the sort advocated by Kaplan (1978a) and McGinn (1981) has now generally given way to salientism about demonstrative reference. Then, I turn to investigate the two most recent,

and promising, versions of salientism on offer: developed by Gauker (2008) and Mount (2008), respectively. I argue that neither of the notions of salience on which each of these proposals relies is ultimately tenable, and thus that each of these views ought to be rejected. In addition, there is reason to suspect that *any* theory of reference built on salience will fail. Specifically, there are cases where it is clear that some object *o* is, by any standard, the most salient object in the context, yet it is equally clear that object *o* is not the referent. On the basis of considerations like these, I urge that salientism be rejected as a viable theory of what fixes semantic values in context.

This leaves intentionalism as the proverbial ‘only game in town’. As was illustrated above, simple versions of intentionalism threaten to collapse into pure Humpty Dumptyism—something which should suffice to make such versions of the view unappealing. However, a variety of more nuanced intentionalist views have been developed in the wake of the Donnellan-MacKay exchange, and many of these views have at least a *prima facie* claim to being genuinely non-Humpty Dumptian. In chapter 3, I introduce a range of more nuanced versions of intentionalism and offer arguments against each of those views. Then I offer my own positive proposal, the constraint theory, and show how it can account for the range of cases that proved either difficult or impossible for its predecessors to handle. What’s more, I claim, there are independent reasons to favor the constraint theory: it is relatively simple, it makes clear predictions, and it is wide in its explanatory scope. In particular, the theory applies as well to names and pronouns as it does to pure demonstratives.

Again, in more detail: chapter 3 begins by introducing the notion of *referential confusion*, which has received a good bit of attention in recent work on epistemology (cf. Millikan 2000, Camp 2002, Lawlor 2005, 2007, Scharp 2005), and then suggests that the classic counterexamples to intentionalism are in fact all productively understood as cases of confusion. This, in turn, suggests that a unified solution to these counterexamples might be possible. I assess the prospects for four amended versions of intentionalism to offer such a solution: Kaplan (1989a) and Siegel (2002)’s ‘perceptual grounding theory’, King (2013a,b)’s ‘coordi-

nation account’, Reimer (1991a, 1992)’s ‘hybrid theory’, and Bach (1992a,b)’s ‘neo-Gricean account’. Reimer (1991a, 1992) and Bach (1992a,b)’s theories, in particular, show some real promise in accounting for the sorts of behavior that demonstratives exhibit when used by speakers who are referentially confused. Unfortunately, both of these theories face some serious drawbacks, both conceptually and in terms of their potential scope. In particular, in spite of being motivated by general considerations regarding referential intentions—and nothing to do with demonstratives specifically—neither theory can accurately account for confused uses of names.

In contrast, my constraint theory not only offers a cleaner account of how demonstrative reference can succeed in cases of confusion, that account can be easily extended to account for a variety of other referential terms. The account begins with the observation that confused speakers are plausibly characterized as having multiple, non-equivalent intentions to refer. It then proceeds to employ the speaker’s choice of one or another referring term to effectively *filter* these intentions for just those that satisfy whatever constraints are conventionally associated with the relevant term. Uses of the name N, for instance, filter speakers’ intentions for just intentions to refer to individuals who actually bear the name N. Similarly, demonstratives filter intentions for relative distance or proximity (depending on whether ‘that’ or ‘this’ has been used). Additionally, when uses of demonstratives are accompanied by ostensive gestures, speakers’ intentions are filtered for just those objects and properties indicated by that gesture. Reference is predicted to succeed when all of the speaker’s various intentions that pass through the relevant filters point to the same object. Otherwise, reference is predicted to fail. As we will see, these predictions are borne out in a wide array of cases involving referential confusion.

Chapter 4 restricts its focus to just neo-Gricean theories of reference, like that of Bach (1992a,b). This sort of theory is also of particular interest both because it is relatively successful and because it is derived, more or less directly, from the broader Gricean theory of meaning—a theory that purports to explain what utterances and their parts mean in

general, not just what referential terms mean. The failure of Bach's theory would thus represent a failure for this wider program of understanding meaning in terms of what Grice calls 'reflexive intentions'. We might therefore hope to learn something important about the nature of meaning more generally from attending to the reasons for this failure.

I argue that Bach's theory fails not just for the reasons outlined in chapter 4, which might seem technical and, thus, ultimately fix-able. In fact, the theory makes implausible predictions in a range of cases involving 'sneaky' speakers.<sup>8</sup> A number of earlier critics pointed to a different sort of sneakiness in order to undermine Grice's claim to have offered a set of necessary and sufficient conditions for an utterance's meaning something on a particular occasion of use (cf. Armstrong 1971, Schiffer 1972, Bennett 1976, Vlach 1981, and Neale 1992). My cases are different in kind from these earlier cases, all of which aimed to call into question only the sufficiency of Grice's proposed conditions for an utterance's meaning something. In contrast, the 'sneaky' cases to which I point aim to undermine the plausibility of Grice's claim to have offered a set of necessary conditions for meaning.

This is significant, in particular, because Grice's critics have continued to accept the necessity half of his proposal, even while they have given up on the claim to sufficiency. Thus, if my arguments are right, it is not just Grice's theory of meaning that is undermined, but any theory that accepts the following, almost pedestrian, claim: that what speakers mean on a given occasion, and what their utterances in turn mean, is determined by their intentions to have particular effects on their listeners. Recall that a claim very much like this one struck Donnellan as so natural that he reached for it without hesitation, and without further argument, in his response to MacKay. What's more, this basic thought underwrites not just Grice (1989d)'s still-popular theory of meaning, but all of its various neo-Gricean successors. Thus, if I succeed in calling this claim into question, this should prompt us to

---

<sup>8</sup>Similar considerations will apply, however, to certain other theories as well. In particular, King (2013a,b)'s 'coordination account' will be susceptible to the arguments developed in this chapter. The common thread between these theories is their insistence that successful reference is grounded in the potential for communicative success.

substantially rethink the very foundations of the theory of meaning.

I conclude in chapter 5 by discussing two possible worries regarding the constraint theory. First, it might seem that intentionalist theories of ‘true’ indexicals—like ‘I’, ‘here’, and ‘now’—are bound to fail. Second, the constraint theory might seem to entail that deceptive uses of language—like the ‘sneaky’ uses of names and demonstratives discussed in chapter 4—are metaphysically on par with non-deceptive uses of language. There are, however, reasons to suspect that a claim to such metaphysical parity would be implausible. *Pace* such worries, I show: first, how the linguistic evidence regarding indexicals fails to tell between non-intentionalist theories and certain sorts of intentionalist theories, of which the constraint theory is one; and, second, that the constraint theory is fully compatible with granting a certain sort of metaphysical priority to cooperative uses of language. Thus, I claim that both of these concerns can be satisfactorily dealt with by the advocate of the constraint theory. Illustrating how will not only help to sharpen our understanding of the constraint theory—it will also help to demonstrate that theory’s prospects for further generalization.

# Chapter 2

## Against Salientism

*In Hypatia the day will come when my only desire will be to leave. I know I must not go down to the harbour then, but climb the citadel's highest pinnacle and wait for a ship to go by up there. But will it ever go by? There is no language without deceit.*  
*Italo Calvino, Invisible Cities—1972*

### 2.1 Introduction

Most philosophers of language nowadays accept that it is some aspect of the speaker's intentions that fixes the meaning, or reference, of demonstratives (e.g. 'this' and 'that'), pronouns (e.g. 'he' and 'she'), and other context-sensitive referring terms in context. Call this view 'intentionalism about reference', or 'intentionalism' for short.<sup>1</sup> A persistent minority, however, claims that intentionalist theories are unable to account for a range of cases involving confused or inattentive speakers. Accounts based on salience, they claim, are better able to explain these cases. On this basis, these 'salientists' advocate giving up on speaker intention-based views and instead adopting the view that it is salience, not intentions, that

---

<sup>1</sup>Cf. Donnellan (1966, 1970), Burge (1974), Kripke (1977), Bertolet (1980), Schiffer (1981), Davies (1982), Kaplan (1989a, 1990), Reimer (1991a, 1992), Bach (1992a,b), Lepore & Ludwig (2000), King (2001, 2013a,b), Siegel (2002) Higginbotham (2002), Fodor & Lepore (2004), Neale (2004), Cappelen & Lepore (2005), King & Stanley (2005), Predelli (2005), Szabó (2006), Soames (2010), and Speaks (2013), among others.

fixes reference in context.<sup>2</sup>

In this chapter, I argue in favor of the consensus view. While salientists present some serious *prima facie* challenges for intentionalist theories of reference, on further investigation most of these challenges turn out to be merely *prima facie*. Moreover, those that remain can in fact be addressed, at least by certain versions of intentionalism. These observations serve to undermine the main argument in favor of salientism—namely, that its primary competitor, intentionalism, suffers from some deep and irreparable flaw. I go on to argue that salientist views of all stripes exhibit some serious drawbacks: even the best salientist views fail to live up to some of their own motivating ambitions, and all varieties of salientism predict, incorrectly, that certain sorts of deceptive uses of language should be impossible. The case for salientism thus looks far weaker than has often been advertised, whereas the reasons for rejecting salientism turn out to be remarkably robust.

Here is the plan for what follows: in section 2.2, I introduce the problem of reference fixing in more detail. Next, in section 2.3, I sketch two recent versions of salientism—advocated by Gauker (2008) and Mount (2008), respectively. I show, further, that by modulating certain aspects of these theories, we can derive a range of other salientist positions. Section 2.4 introduces the main salientist arguments against intentionalism. In section 2.5, I examine these arguments and conclude that they are not much good. Some of these arguments invoke criteria that salientists themselves cannot meet, whereas others are based on misunderstanding the range of intentionalist views actually available. Section 2.6 introduces several problem cases for salientist theories that no variant of the view looks well-equipped to handle. On the other hand, intention-based theories are better-situated to handle each of these cases. In section 2.7, I conclude from all this that we have strong reason to favor intentionalist

---

<sup>2</sup>Cf. Wettstein (1984), Reimer (1991b), Gauker (2008), and Mount (2008). There is also a related tradition (particularly popular in linguistics), according to which demonstratives and pronouns are anaphoric on referents that are already available in the common ground of the conversation. More specifically, these terms are generally supposed to be anaphoric on the referent which is (somehow) designated as most salient in the common ground (cf. Isard 1975, Lewis 1979, Heim 1982, Clark et al. 1983, and Roberts 2002.). While I will not specifically address this tradition here, all of the arguments below should carry over *mutatis mutandis*.

theories of reference as opposed to salience-based ones.

## 2.2 The Problem of Reference Fixing

As outlined in chapter 1, in order to pursue the project of truth-conditional semantics in earnest, we need some way of associating uses of context-sensitive terms with semantic values in context. This problem arises for a variety of different sorts of terms, such as quantifiers, definite descriptions, certain types of property terms (e.g. colors, comparative adjectives), names (once we relax the assumption that there is only one person named N per speech community), pronouns, and demonstratives. If we cannot associate such terms with values in context, then we would be unable to associate sentences containing such terms with truth-conditions, relative to a particular context.<sup>3</sup>

In this chapter, we will restrict our attention to what fixes the values—or, more specifically, the reference—of demonstratives in context.<sup>4</sup> The reason for this restriction is simple: these are the terms on which advocates of salientism have primarily focused, and the terms that they have used to motivate their view as against intentionalism. That said, it is worth

---

<sup>3</sup>With regard to some of these—e.g. quantifiers, colors—we might, instead of trying to solve the problem of value fixing, try being invariantists. That is, we might endorse the claim that these terms are not, in fact, context-sensitive at all. This will mean that many of the things that we say thinking them to be true are, in fact, false. But perhaps this is a tolerable price to pay for a simple, streamlined semantic theory. I take it that invariantist theories of names, pronouns, and demonstratives, however, are more difficult to countenance. Since we will focus here only on instances of the latter sort of term, I will not attempt to motivate a variantist approach to each of these.

<sup>4</sup>Traditionally, demonstratives have been assumed to denote objects relative to a context. Recently, however, this claim has been challenged by a number of philosophers claiming that demonstratives are best thought of as denoting complex quantifiers in context (cf. Taylor 1980, Neale 1993, 2004, King 1999, 2001, 2008, and Lepore & Ludwig 2000). Still, these ‘quantificationalists’ will face an analogue of the reference fixing problem, at least for unbound uses of demonstratives (the only type of use we will consider here). This follows from their claim that *which* complex quantifier a use of a demonstrative will denote in context is determined by a particular, designated object—which itself must be determined by the context somehow (most quantificationalists have assumed that the aspect of the context that fulfills this function is the speaker’s intentions, but nothing about quantificationalism requires this). Quantificationalists are thus invited to conceive of the dialectic in the main text as concerning what fixes the semantic values of demonstratives in context, rather than their reference specifically (since, according to quantificationalists, these aren’t referential terms). Similar considerations will apply in subsequent chapters regarding the use of names, which some have argued are in fact devices of predication rather than reference (cf. Burge 1973, Bach 1991, 2002, Geurts 1997, 1999, Elbourne 2005, and Fara 2011, 2013b).

noting that one might at least hope for a unified account of how the semantic values of context-sensitive terms in general are determined in context. What's more, appeals to either salience or to speakers' intentions look to have the potential to generalize to terms beyond just demonstratives; in fact, that would seem to be a welcome feature of both such views. If, as I argue, salience-based theories fail *even for* demonstratives, then it is probably not worth trying to extend such theories to other domains. In subsequent chapters, however, I will be concerned with developing an intentionalist theory of reference capable of accounting for a large class of context-sensitive terms, not just demonstratives.

Before moving on, it will help to briefly clarify what exactly the reference of demonstratives is supposed to explain. The notion of reference is standardly invoked in order to help answer two inter-related questions about sentences containing demonstratives: first, what are such sentences *about*, in an intuitive sense, in context? And, second, under what conditions would utterances of those sentences be *true* or *false*? Naively, we might expect for these two issues to be related, with sentences being true in context when they attribute properties to objects (the objects those sentences are about, specifically) that those objects actually bear, and false when they attribute properties to objects that those objects do not actually bear. Below, we will largely be focused on this latter question. However, it will often help to bear this naive (and, I think, rather plausible) picture of reference in mind when trying to understand some of the motivations for salientism.

## 2.3 Salientism

Salientism can be legitimately regarded as the successor to an earlier theory of demonstrative reference, one advocated at different points by both Kaplan (1978a, 1989b) and McGinn (1981). According to that theory—call it ‘demonstrationism’—demonstrative reference is determined by the ostensive gestures that very often accompanies uses of demonstratives.<sup>5</sup>

---

<sup>5</sup>In fact, the basic demonstrationist idea dates back even further—at least to Quine (1968). My own proposed version of intentionalism, the constraint theory, picks up on certain aspects of the demonstrationist

For McGinn, for instance, “the referent of a token of ‘that F’ is to be the first F to intersect the line projected from the pointing finger, i.e. the F at the place indicated—one might almost say geometrically—by the accompanying gesture” (McGinn 1981, p. 163). Similarly, for Kaplan, “[d]emonstratives are incomplete expressions which must be completed by a demonstration” (Kaplan 1989b, p. 527). It is whatever object that demonstration picks out in context which serves as the referent of the relevant demonstrative.

Problems with this theory are rife. First, and most obviously, many uses of demonstratives are unaccompanied by any sort of ostensive gesture or physical demonstration. Contrary to what the demonstrationist is bound to predict, it is not the case that all such uses result in reference failure; in fact, many of these uses are cases of referential success. Second, we seem to be capable of referring to more than just objects with uses of demonstratives; we can also refer to their surface features, other properties that they bear, and even things that they represent. Demonstrationism, at least in its unmodified form, is unable to account for this fact. Third, it is by no means trivial to determine just what is demonstrated by an ostensive gesture. Even granting ourselves McGinn’s apparatus, for instance, we would require some method of determining *which* ray extending from the outstretched finger counts as ‘the line of the pointed finger’. What’s more, we would require a similar rubric for determining the relevant ray with regard to other methods of pointing as well—such as pursing one’s lips.<sup>6</sup>

One might, of course, try to salvage demonstrationism in light of such challenges. As it happens, however, few have been tempted to attempt such a rehabilitation.<sup>7</sup> Rather, as

---

idea as well. This will become clearer in chapter 3.

<sup>6</sup>Lip pursing is actually a rather common method of pointing in non-Western languages. For discussion, see Key (1962), Sherzer (1972, 1983, 1993), Poyates (1983), Feldman (1986), Enfield (2001), Wilkins (2003).

<sup>7</sup>Two partial exceptions might seem to be Roberts (2002) and Reimer (1991a, 1992). These theories are complicated, however, and I doubt that either is ultimately best thought of as a version of demonstrationism. On the surface, Roberts (2002) seems to propose a sort of demonstrationist-salientist hybrid, according to which demonstrating something suffices to make it maximally salient. However, Roberts’ notion of a demonstration itself is plausibly an intentionalist one, not a purely gestural one like Kaplan’s or McGinn’s. This makes Roberts’ theory somewhat difficult to classify; I suspect that it is ultimately best conceived of as a sophisticated version of salientism, though I will not argue for that at any length here. Reimer (1991a, 1992), on the other hand, defends what she calls a ‘quasi-intentionalist’ view, according to which it is the speaker’s intention to demonstrate a particular object that determines reference. According to Reimer, these

indicated at the outset, most philosophers have instead followed Kaplan's own change of heart and adopted one or another version of intentionalism.<sup>8</sup> By contrast, salientists have been tempted by something like the following line of thought: the reason that demonstrationism failed is not because it was fundamentally on the wrong track. Rather, the problem is that demonstrationists were overly concerned with ostensive gestures. Gesturing at something, however, is merely *one way* of making it salient in context. It is salience that fixes reference in context, not ostension *per se*. Essentially, the claim is that demonstrationists mistook the mechanism of salience-raising for its ultimate result—that is, reference. Salientists, on the other hand, have recognized that gesture is merely a one way of manipulating salience; it is salience itself that actually fixes reference.

Having come this far, a new question arises: what exactly is salience? Here, we find significant divergence across different theorists. Basically, there are two axes of disagreement regarding what salience amounts to—or, more precisely, about which notion of salience is relevant for the task of explaining linguistic reference. The first axis has to do with the relevant *perspective* for salience: is it salience for the speaker that matters, salience for the listener, or both? Or is it rather some sort of notion of objective salience, salience from no perspective in particular? The second axis has to do with what sorts of *features* are relevant for determining salience: is it merely visual or sensory salience? Or can something be salient for a variety of other reasons as well, like having been mentioned previously in conversation, or being somehow related to something that is visually salient?

Two of the most nuanced, proponents of salientism have gone very different ways with

---

intentions will themselves often reference gestures—thus imbuing those gestures with semantic significance. However, it is still fairly clear that, on Reimer's view, it is ultimately intentions to refer to things gestured at, rather than physical demonstrations themselves, that serve to fix reference.

<sup>8</sup>Kaplan characterizes his change of heart nicely as follows: "I am now inclined to regard the directing intention, at least in the case of perceptual demonstratives, as criterial, and to regard the demonstration as a mere externalization of this inner intention. The externalization is an aid to communication, like speaking more slowly and loudly, but is of no semantic significance" (Kaplan 1989a, p. 582). While the reasons for Kaplan's shifted allegiance are never fully explicated in Kaplan (1989a), one suspects that Kaplan may have realized that determining which object or property a particular physical demonstration picks out will not be as trivial as McGinn (1981) makes it out to be, nor as trivial as Kaplan himself had previously taken it to be. Kaplan (p.c.) has in fact confirmed this suspicion.

regard to the first of these options. On the one hand, Mount (2008) has endorsed a view on which what counts as salient for the purposes of reference fixing is whatever object is “mutually recognized [by both the speaker and listener] as maximally salient” (p. 154). This entails *inter alia* that the salient object in a context is the object that is maximally salient to each of the conversational participants—and hence, according to Mount, the “focus of [their] perceptual or cognitive attention” (Mount 2008, p. 154). Mount’s preferred notion of salience is thus explicitly mind-dependent, and, as she puts it, ‘non-objective’.<sup>9</sup> Gauker (2008), on the other hand, appeals to an ‘objective’ notion of salience in order to fix reference. According to Gauker, both the speaker and listener can actually be *mistaken* about what is salient in context (Gauker 2008, pp. 364–66).<sup>10</sup> Mount and Gauker agree, however, with regard to the second axis: it isn’t just perceptual salience that matters. Rather, it is salience in some wider sense—one that incorporates considerations like what is in the conversational record or the common ground—that matters for fixing linguistic reference.<sup>11</sup> In spite of this point of agreement, Mount and Gauker’s respective versions of salientism differ substantially from one another. Consideration of both will prove helpful in the following respect: between the two theories, we will be able to see why the problems I raise below are not just problems for these particular versions of salientism, but rather ones that will arise for pretty much any of the ways in which salientism might conceivably be developed.

---

<sup>9</sup>I very much doubt that we should endorse the claim that mind-dependence entails non-objectivity, as Mount (2008) seems to assume. However, I leave that issue aside for present purposes.

<sup>10</sup>Note that Gauker (2008) doesn’t explicitly consider his view a salience view, since he takes it that salience is restricted to purely perceptual features. However, since most others who use the term disagree with him on this score, it strikes me as fair to characterize his view as falling within the salientist tradition, broadly understood.

<sup>11</sup>Historically, Clark et al. (1983) and Wettstein (1984) endorse a picture much like Gauker’s. Others have tended to be less specific about what they mean by ‘salience’. One exception is Reimer (1991b), in what appears to be her earliest work on the subject. There, Reimer endorses a speaker-oriented version of salientism—on which the speaker refers to whatever she happens to be *attending* to at the moment of utterance. I will not address this latter view at length in the main text since it strikes me as fairly clear that speakers can use demonstratives to refer to objects that are not the focus of their cognitive attention at the time of utterance. Suppose, for example, that a woolly mammoth is chasing me and I know that there is a spear next to you. I might refer to that spear in yelling “Throw that spear at the mammoth!” Plausibly, however, either the mammoth itself (is it gaining on me?) or the ground in front of me (are there hiding places?) will be the focus of my cognitive attention in these circumstances, not the spear.

## 2.4 Arguments for Salientism

Now I turn to the task of introducing the main arguments for salientism. Two arguments have emerged in the literature: one from cases and the other from what listeners can reasonably be expected to recover in context. Both of these arguments aim to show that intentionalism is false—and thus that, as the only viable alternative, salientism must be true.

### 2.4.1 The Argument from Cases

The most persistent argument for salientism runs as follows: there are cases in which the speaker intends to refer to some object  $o$  with her use of a demonstrative, but  $o$  is not the referent. Thus, intentionalism must be false and salientism is vindicated.

Consider, for instance, a case introduced in Wettstein (1984): Smith is in the distance raking leaves, but the speaker mistakes him for Jones. Pointing at this man (i.e. Smith), the speaker utters the sentence “That is a self-destructive man. He has been raking leaves against his doctor’s orders” (Wettstein 1984, p. 70).<sup>12</sup> According to Wettstein, the person who the speaker ‘has in mind in the primary sense’ is Jones. Thus, if intentionalism is correct, it is Jones, Wettstein reasons, who should turn out to be the reference of ‘that man’ in this context. But, in fact, the referent here is clearly Smith (Ibid., p. 71).

In a similar vein, Mount introduces a variant of Kaplan (1978a)’s classic ‘Carnap-Agnew’ case. In the original case, the speaker, without looking, points directly behind herself—and directly at a picture of Spiro Agnew—and utters the sentence:

- (1) That is a picture of one of the greatest philosophers of the twentieth century.<sup>13</sup>

The speaker, however, is under the mistaken belief that behind her hangs a picture of Rudolf Carnap. Kaplan’s reaction was to think that the speaker has uttered a falsehood in this case,

---

<sup>12</sup>The case is clearly a variation on Kripke (1977)’s classic Smith-Jones case, which involves names rather than demonstratives. For further discussion, see chapter 3.

<sup>13</sup>While the example is basically Kaplan (1978a, p. 239)’s (22), I have taken the liberty of substituting the ordinary English ‘that’ for Kaplan’s use of his technical ‘dthat’-operator.

in virtue of both the fact that her use of ‘that’ referred to the picture of Agnew, rather than to the picture of Carnap, and the fact that Agnew was not one of the greatest philosophers of the twentieth century.

This case already presents a *prima facie* problem for intentionalism, since it seems plausible that, in some natural sense, the speaker intended to refer to the picture of Carnap. Mount then modifies the case as follows: suppose that both the speaker and listener are facing away from the picture of Agnew, and suppose further that both believe the picture hanging behind them to be a particular picture of Carnap. The speaker once more points behind herself and utters (1). Mount claims that, in contrast to the more standard version of the Carnap-Agnew case, the speaker here *succeeds* in referring to the picture of Carnap—in virtue of her having succeeded in making that picture uniquely salient, relative to both the speaker and the listener’s mistaken beliefs about the context. Initially, this case might look less problematic for the intentionalist than the original; after all, the speaker presumably intended to refer to the picture of Carnap. As we’ll see below, however, things are more complicated than this: many of the versions of intentionalism that are capable of handling the original Carnap-Agnew case will predict that, here too, the speaker has referred to the picture of Angew. Mount’s case is thus aimed at forcing intentionalists into a dilemma: account for the original Carnap-Agnew case or this modified case, but not both.

Finally, Gauker (2008) offers the following case:

Harry has wrapped a garish pink-and-green tie around his neck and is looking at himself in a mirror. Sally is standing next to the mirror gazing toward the tie around Harry’s neck and says, “That matches your new jacket.” As a matter of fact, Sally has been contemplating in thought the tie that Harry tried on two ties back...We can even suppose that in saying ‘that’ what she intended to refer to was the tie two ties back. (Gauker 2008, p. 363)

Gauker claims that, in these circumstances, Sally’s use of ‘that’ refers to the tie that Harry is currently wearing, rather than to the one that Sally is thinking about and intending to refer to. This case, Gauker reasons, thus constitutes a straightforward counterexample to intentionalism (Ibid., p. 363).

## 2.4.2 The Argument from Recoverability

Gauker (2008) has recently supplemented the argument from cases with a separate argument that runs as follows:

Hearers have little access to what people have in mind apart from the interpretation of what they say. So interpretation would be a problem that hearers could not solve if they had to know what a speaker intended in order to identify the content of the context and thus to interpret the speaker's utterance. So a hearer cannot reliably employ a method of interpretation that requires the hearer to have an independent insight into the speaker's intention. But, as we have seen, the proposition expressed by an utterance has to be one that a competent hearer could assign to it using a method of interpretation that the hearer could reliably employ on the basis of features of the situation that the hearer could normally be aware of. (Gauker 2008, p. 362)

As I understand it, the argument amounts to this. Semantic theories aims to associate utterances of sentences with propositions (the propositions 'expressed' by these utterances). Those propositions, in turn, must be recoverable by a competent hearer via some reliable method of interpretation. But since listeners cannot reliably discern what others have in mind, what the speaker has in mind cannot constitute even a partial determinant of the proposition expressed by a particular utterance.

We might, however, wonder what support there is for Gauker's claim that propositions expressed must be recoverable by listeners in this fashion. Here is Gauker's explanation,

Semantic theory...stands in the service of a theory of linguistic communication according to which communication takes place (at some shallow level at least) when the hearer interprets the speaker's utterance as expressing the proposition that it does in fact express. If the account of expressing a proposition that our semantic theory provides is to serve a theory of linguistic communication in this way, then the manner of interpretation that discovers the proposition expressed must be one that a competent hearer could reliably employ on the basis of features of the situation that the hearer could normally be aware of. So the proposition an utterance expresses will be a proposition that a competent hearer would be able to assign to it by means of a method of interpretation that the hearer could reliably employ on the basis of features of the situation that the hearer could normally be aware of. (Gauker 2008, p. 361)<sup>14</sup>

---

<sup>14</sup>Gauker (2008) originally introduces this 'recoverability' criterion for semantic content in the context of

In other words, semantic theory is constrained by the requirement that it play a productive role in a relatively simple theory of communication, according to which the semantic content of an utterance of S *just is* the content that a speaker could reasonably hope for a listener to recover on the basis of that utterance. This, in turn, means that semantic content cannot itself be determined by factors that listeners are in no position to reliably discern in context. If Gauker is right about our ability to access each others' intentions, then intentions cannot play a role in fixing semantic content in context. Saliency, on the other hand, *is* something that listeners are reliably able to identify in context—according to Gauker at least. Thus, saliency is the sort of thing we might reasonably appeal to in offering a theory of what fixes reference in context. Again, this would seem to favor salientism as against intentionalism.

## 2.5 Why These Arguments Fail

In this section, I explain why neither of these arguments succeeds. Basically, the argument from cases fails because the cases themselves either don't show what they purport to show—or else because they rely on some highly dubitable judgments. Gauker's argument from recoverability, on the other hand, assumes that saliency is far more accessible, and speaker's intentions far less accessible, than we should be prepared to grant. What's more, it rests on a dubious assumption about the aim of semantic theory.

---

having supposed, for the sake of argument, that semantic theory does indeed stand in the service of a theory of communication. However, Gauker later relies on some intermediate conclusions that are supposed to follow from this picture in the course of offering his argument against intentionalism (as in the first block quote in main text above). Since Gauker never argues that intentionalists are either already implicitly committed to, or else must for some reason accept this picture of how semantics and the theory of communication relate to each other, it seems that either: (i) Gauker's argument is fallacious—or, more specifically, question-begging—since it relies on a premise granted solely for the sake of argument well after the supposition that this premise is correct has in fact expired (the above block quotes in the main text are from section 2 and section 1 of Gauker's paper, respectively); (ii) Gauker's conclusion is meant to be merely conditional on accepting this picture of how semantic theory and the theory of communication are supposed to relate to each other, despite the fact that he offers no indication that he is offering only a conditional conclusion; or (iii) Gauker takes this particular picture of the relationship between semantic theory and the theory of communication to be so compelling that, once it is introduced, it stands in no need of further justification. I take it that (i) is relatively uncharitable and strive in the main text to remain indifferent between whether (ii) or (iii) is the proper interpretation of Gauker (2008). I return to this issue in subsection 2.5.2.

### 2.5.1 Why the Argument from Cases Fails

First, consider Wettstein’s variant of the Smith-Jones case. Wettstein claims that the speaker ‘primarily has Jones in mind’, and that intentionalists are thus committed to Jones being the referent. But it is perfectly natural to think that the speaker also intends to refer to the man in the distance who she both perceives and mistakes for Jones (i.e. Smith). So it should be open to intentionalists to claim that it is *this* intention which is relevant for fixing reference, not the speaker’s intention to refer to Jones. An intentionalist inclined to go down this route would, of course, need to offer some explanation for why it is the intention to refer to Smith, rather than the intention to refer to Jones, that matters in this context. But, in fact, a number of different ways of explaining why it is the Smith-intention that matters have already been proposed (cf. Schiffer 1981, Kaplan 1989a, Reimer 1991a, 1992, Bach 1992a,b, Siegel 2002, and King 2013b).<sup>15</sup> For example, according to Kaplan (1989a), it is only intentions grounded in occurrent perception of an object—that is, perception of that object at the time of utterance—that can serve to fix reference. In Wettstein’s case, the speaker is only occurrently perceiving Smith. According to Reimer (1991a, 1992), on the other hand, it is the speaker’s intention to refer to an object of a particular sort (here, a man) in the direction of her pointing—an intention specified *de dicto*—that serves to fix reference, *not* her intention to refer to one or another object specified *de re*. Adopting either of these options would suffice to answer Wettstein’s challenge.

Turning now to Mount’s version of the argument, the first thing to note is that, once more, quite a bit hinges on *just which* version of intentionalism we are considering as her target. Kaplan’s original ‘Carnap-Agnew’ case has long been known to cause trouble for intentionalists. Accordingly, a number of versions of intentionalism have been developed with just that case in mind. Kaplan (1989a)’s own version of intentionalism, for instance, is explicitly restricted to cases in which the speaker is occurrently perceiving her intended

---

<sup>15</sup>My own preferred version of intentionalism, outlined in chapter 3, can also account for this judgment. Some recent intentionalists (e.g. King 2013a) might instead be tempted to dispute this judgment.

referent. Hence, the Carnap-Agnew case falls outside of the domain of the theory, allowing Kaplan to remain explicitly non-committal about how such cases should be handled. Or consider Reimer (1991a, 1992)'s proposal: it is the speaker's intention to refer to the picture in the direction of her pointing that serves to fix reference. This means that she will have referred to the picture of Agnew, not the picture of Carnap, with her use of 'that'—since only the picture of Agnew lies in the direction of her pointing.

Now consider Mount's variant of the Carnap-Agnew case, where neither the speaker nor listener is facing the picture of Agnew, and where both believe it to be a picture of Carnap. Kaplan can respond just as before: his version of intentionalism cannot be threatened by the case, since it makes no predictions about it. Still, this may not strike the reader as a very satisfactory response. So consider instead a more ambitious version of intentionalism, along the lines suggested by Reimer (1991a, 1992). Reimer predicts that the speaker's use of 'that' refers to the picture of Agnew at which she actually points, rather than to the picture of Carnap. This is not the result that Mount thinks we should be aiming for. But I think we should ask in all seriousness: is Mount's judgment regarding the case actually correct?

It will help to consider a similar case, one in which the speaker and listener are both facing the picture of Agnew but are standing some distance away from it. Because of this distance, neither the speaker nor the listener has noticed that the picture they are seeing is a picture of Agnew, rather than the picture of Carnap that usually hangs in this location. In other words, both mistake the picture of Agnew they are occurrently perceiving for the picture of Carnap with which they are already acquainted, and which they expect to be hanging in this location. As in Mount's version of the case, it seems that when the speaker points and says (1), she is likely to succeed in transmitting a thought about the picture of Carnap—the picture that both she and the listener mistakenly take themselves to be looking at. At the very least, the listener is likely to draw inferences like: the speaker believes that Carnap is one of the greatest philosophers of the twentieth century. Yet here, I take it, it is quite clear that the speaker's utterance of (1) is false, despite the role that utterance plays

in communicating something true.

Similarly, it seems to me quite plausible that Mount's judgment regarding her modified version of the Carnap-Agnew case is simply incorrect: although the speaker's utterance was false, due to the odd circumstances under which it took place, it served to communicate something true. What's more, it is worth noting that I am not alone in thinking something like this; similar judgments are attested to in Kaplan (1978a), Bach (1992b), and Reimer (1992). Thus, it seems to me that there is ample reason to be skeptical of Mount's judgment that the speaker's utterance of (1) is true in her version of the case. But if this is right, then Mount's case serves to undermine *no* version of intentionalism.

Finally, let us consider Gauker's necktie case. Gauker claims that when Sally intends to refer to the necktie that Harry tried on two ties ago, her use of 'that' nonetheless refers to the tie he has on *now*. As it stands, however, it seems to me that Gauker's case is significantly under-described: has Sally noticed the switch? Has she pointed? Does she say anything more than just this one sentence?<sup>16</sup> Suppose, for instance, that Sally were to follow up on her initial utterance of, "That matches your new jacket," with "But I don't like the polka dots." Suppose further that the tie that Harry had on two ties back had polka dots, but that the one he presently has on does not. In this instance, it is not at all clear that Harry would be unable to discern which tie Sally has in mind. Nor does it seem right to say that Sally has referred to the garish pink-and-green tie that Harry is currently wearing; on the contrary, it strikes me that, plausibly, Sally has said something true about the tie Harry had on two ties ago.<sup>17</sup> If that's right, then either the reference of Sally's use of 'that' somehow shifts when this second sentence is added—meaning that what terms refer to can actually change over time, and depends at least in part on events subsequent to their utterance—or else Sally was referring to the tie two ties back all along—even though this fact is unlikely to have been evident to the listener without the addition of the second sentence. The latter

---

<sup>16</sup>Whether or not Sally has noticed the switch, for instance, is critical to what neo-Gricean theories, like Bach (1992a,b)'s, will predict in this case. For explanation why, see chapter 3 and chapter 4.

<sup>17</sup>Thanks to David Kaplan for suggesting this example.

explanation, it seems to me, is both simpler and more plausible than the former. Accepting that explanation, however, requires the rejection of Gauker’s judgment on the Harry-Sally case. In section 2.6, I will introduce a related case—one which points even more clearly towards the conclusion that we should reject Gauker’s judgment here.

Before turning to that, however, it is worth noting that, once more, intentionalists are by no means required to claim that Sally’s use of ‘that’ referred to the tie two ties back. Since Sally is only occurrently perceiving the garish tie that Harry is now wearing, Kaplan (1989a)’s theory will predict, with Gauker, that Sally’s use of ‘that’ refers to the tie that Harry is now wearing. And supposing that Sally has pointed or gestured in some other way, then Reimer (1991a, 1992)’s theory will match Gauker’s judgment as well. That’s because, according to Reimer, Sally’s only intention that matters for fixing reference is her intention to refer to the tie in the general direction in which she gestures.

What we have seen is that at least one version of intentionalism, i.e. Kaplan (1989a)’s, is compatible with the salientists’ judgments on all of these cases, while another, i.e. Reimer (1991a, 1992)’s, is compatible with two of the three. Some may find Kaplan (1989a)’s theory unsatisfying, given its limited scope, but that objection is independent from the argument from cases.<sup>18</sup> This, however, is not the only reason to doubt the efficacy of the argument from cases: the clearest of the three case to which salientists appeal, i.e. Wettstein’s, is in fact fully compatible with most extant forms of intentionalism.<sup>19</sup> The other two cases, i.e. Gauker’s and, in particular, Mount’s, are incompatible with a wider range of intentionalist theories. However, the force of these cases also relies on some highly dubitable judgments. Thus, while it is no doubt an argument worth taking seriously, the argument from cases looks to be far from conclusive.

---

<sup>18</sup>In fact, this is an argument that I myself will run against Kaplan’s theory in chapter 3.

<sup>19</sup>As noted above, Wettstein’s case is compatible with at least the versions of intentionalism put forward in Schiffer (1981), Bach (1992a,b), Siegel (2002), King (2013b), and Speaks (2013), in addition to Kaplan and Reimer’s versions. The case is also fully compatible with my own ‘constraint theory’, introduced in chapter 3. In the course of that chapter, I also introduce Bach and Siegel’s views in significantly more detail.

## 2.5.2 Why the Argument from Recoverability Fails

This leaves Gauker’s argument from recoverability. Recall that, according to Gauker, semantic content cannot be determined, even in part, by aspects of the context that listeners cannot reliably assess. Were semantic content to be so determined, Gauker reasons, “interpretation would be a problem that listeners cannot solve” (Gauker 2008, p. 362). And this result is something that cannot be tolerated.

Unfortunately, Gauker never elaborates on this claim. Nor does he give any substantive reasons in favor of it. What he does tell us, early on in his paper, is that “[a]ccording to a common picture, linguistic communication is a matter of a speaker’s expressing propositions by means of utterances and a hearer’s understanding which proposition the utterance expressed” (Ibid., p. 360). Here is one way of trying to understand what Gauker’s worry might be: suppose that this common picture of communication is correct, and suppose further that communication regularly succeeds when speakers use demonstratives. Then demonstrative reference must be determined by some aspect of the context that listeners can reliably assess—that is, by some *public* aspect of the context. Were demonstrative reference determined, even in part, by some aspect of the context that listeners cannot reliably assess, then these listeners would not be able to recover the referents of demonstratives nearly as regularly as they actually do. In other words, we would expect to see a far higher rate of communicative failure than we actually observe. On this way of understanding Gauker, it is this incongruence between predicted and observed failure rates that cannot be tolerated.<sup>20</sup>

In response to Gauker, the first thing to note is that, for all that Gauker has actually said, it is open to the intentionalist to simply deny the picture of communication that is required for the argument to work. That is, if we are interested primarily in the project of associating utterances with truth-conditions, it is not immediately obvious why these truth-conditions need bear any direct relationship to conditions of communicative success.

---

<sup>20</sup>Note that Gauker (2008) offers no evidence regarding how reliable communication using demonstratives actually is. Rather, he simply assumes that it is highly reliable. This strikes me as dubitable at the very least. For present purposes, however, I set this worry to the side.

After all, we might be inclined to think that communication is achieved in instances where the listener recovers what is *conversationally implicated* by an utterance—or more generally, where he recovers what the *speaker means* by an utterance—even though that listener fails to recover the actual truth-conditions of the utterance.

Consider Grice’s example “There is a garage around the corner” as said to a driver clearly in need of gasoline (Grice 1989a, p. 32). On hearing this, the listener may immediately think “Ah, I can purchase some petrol around the corner.” And, indeed, this seems to be both what the speaker meant and what she implicated. According to Grice, this must be the sort of thing that a listener could, in principle, work out via reasoning about the rules of conversation (embodied in Grice’s proposed conversational maxims) and the strict meaning of the sentence uttered. But in practice, Grice concedes, we may well just ‘intuitively grasp’ what was meant or implicated, without actually going through any such chain of reasoning (Ibid., p. 31). Likewise, we can imagine communication succeeding in an instance where the listener recovers what the speaker intended to convey with a bit of loose speech, even though he fails to recognize what the utterance meant, strictly speaking. Gauker’s assumption about the relationship between semantics and communication thus looks to be decidedly non-trivial. But unless it holds, the argument from recoverability is unsound.

For the sake of argument, let us simply grant Gauker the required premise regarding the relationship between truth-conditional content and communication. Now, let us consider in more detail how Gauker himself takes the reference of demonstratives to be fixed in context. According to Gauker, reference is determined by a ‘best fit judgment’ regarding at least six *accessibility criteria*—or, in our terminology, criteria of salience. For Gauker, what a speaker refers to in context is the object that, all things considered, is the most perceptually salient, the most salient in prior discourse, the most relevant, the most clearly gestured at, the most sensible in terms of fitting in a series, and, lastly, is most charitably understood as the speaker’s intended referent (Gauker 2008, pp. 364–66). This, in more detail, is what Gauker conceives of as constituting the property of being maximally salient in a context. In

line with his own desideratum, each of these criteria is plausibly publicly accessible.<sup>21</sup>

The question then is whether the property of being the object that the speaker intends to refer to might also count as publicly accessible. Gauker (2008) assumes that the case against this is clear, but he offers no real argument to that end. In contrast to Gauker, it strikes me that the case for the public inaccessibility of speakers' intentions is far from clear. Consider that, as both Reimer (1991a, 1992) and Bach (1992a,b) have stressed, speakers typically go to some lengths to make their referential intentions evident to their listeners: they gesture, they choose particular referential terms over others, they use particular predicates, they rely on what they take to be visually salient or striking at the context, at what has been under discussion in the conversation, and on what they know about the listener's background beliefs and interests. So while speakers' intentions certainly are publicly inaccessible *in principle*—that is, they are the sorts of things that certainly *can be* inaccessible to listeners—*in practice* this is likely to be the case only quite rarely.<sup>22</sup> While speakers are hardly perfect at making their intentions evident to their listeners, it is not at all implausible to think that they are reliable enough at doing so to explain the sorts of regularities in communicative success to which Gauker points. Thus, it seems that intentionalists can meet Gauker's challenge on its own terms: there is ample reason to think that speakers, aiming to be understood, will go to some lengths to make their intentions public. If this is right however, then Gauker's argument fails to favor salientism over intentionalism.

## 2.6 Why Salientism Should be Rejected

So far, I have focused primarily on what salientism amounts to, what the motivations for it are supposed to be, and why these motivations aren't much good. Now, I turn to the task of explaining why salientism should in fact be rejected.

---

<sup>21</sup>While I have my doubts that these criteria are in fact well-defined, I leave aside that worry for the time-being. More on this in section 2.6.

<sup>22</sup>In fact, for Bach (1992a), intending to make one's referential intentions public is in fact *required* for those intentions to succeed in fixing reference (p. 144).

Salientism is often dismissed in passing for being too ill-defined to constitute a satisfactory theory, let alone one that can be used to generate actual predictions. But, as we have seen, this charge won't stick against at least some contemporary versions of salientism. Mount (2008), for instance, claims that being salient to an individual amounts to being “the focus of [her] cognitive or perceptual attention” (p. 154). Being mutually maximally salient, then, is to be the focus of cognitive or perceptual attention for all parties to the conversation. While we might hope for some further elaboration on what it means to be ‘the maximum point of cognitive or perceptual attention’, there is nothing deeply opaque about this overall picture. Nor is there anything particularly opaque about Gauker’s claim that referents are those objects which, in context, best fit his six criteria of objective salience. I therefore think it best to leave aside this potential line of criticism.

Instead, I will argue for three things: first, Mount’s notion of mutual maximum salience entails, implausibly, that reference fails whenever listeners, for whatever reason, happen not to be paying attention. Second, there is reason to suspect that Gauker’s notion of objective salience is in fact conceptually incoherent. In particular, his six criteria of salience are all plausibly ‘perspective-relative’ (in a sense to be explained below), and thus cannot be objective in the way that he needs them to be. And, third, even ignoring the preceding worries, salientism of any stripe is bound to get certain cases wrong—including cases involving speakers whose explicit aim is to deceive their listeners about what they are saying.

Let us start with consideration of Mount’s suggestion that what fixes reference is being the point of cognitive focus for both the speaker and listener. Note that this criterion for reference entails that almost any time that a listener happens not to be paying sufficient attention to the speaker, reference fails. This follows from the fact that, in such cases, the listener will typically be attending to some object other than what the speaker is attending to. Thus, there will be no single object that is the focus of attention for both the speaker and the listener—unless, that is, the listener *just happens* to be attending to the object to which the speaker is also attending. I submit that it hardly seems right to think that

whenever a listener fails to properly attend to what the speaker is saying, and the speaker happens to use a demonstrative, the listener's inattentiveness alters the truth-value of the speaker's utterance. This, however, is just what Mount's view entails—since reference failure is precisely the right sort of thing to make an utterance true, false, or even truth-valueless.<sup>23</sup>

Next, consider Gauker's claim that there will generally be a single object that best fits his six criteria of salience in context, all things considered. Gauker argues at some length for the legitimacy of invoking all-things-considered judgments in philosophy, and I have no wish to dispute the general legitimacy of appeals to such judgments here (Gauker 2008, pp. 366–67). Rather, my concern has to do with certain specific aspects of Gauker's theory: while there might sometimes, or even often, be widespread agreement about what best satisfies the six criteria of salience that Gauker identifies, there is reason to doubt that such agreement reflects any underlying fact of the matter in these cases. For example, consider a context that contains, among other things, a rampaging woolly mammoth. It will initially seem plausible, I take it, to claim that the mammoth is the most visually salient thing in the context. But note that if I am desperately searching for a good hiding place, then the hollow under a large tree may be far more visually salient to me, in context, than the mammoth. What this scenario illustrates is that visual salience is sensitive not just to which objects, with what properties, are to be found in a context, but also to the observer's 'perspective' on that context—that is, to her interests, beliefs, values, emotions, prior attentional states, visual

---

<sup>23</sup>Mount (2008), being aware of this potential worry, offers the following response: “if the misunderstanding occurs because the hearer is not paying attention or is not a full conversational participant (given whatever independently-motivated theory we have for conversational participation), then there can be misunderstanding of reference on the hearer's part without there necessarily being failure of reference on the speaker's part” (p. 156). The problem with this response is that it isn't clear what this notion of 'reference on the speaker's part' amounts to on Mount's theory, nor does Mount elaborate further. If Mount just means *speaker reference* by this—that is, something like 'what the speaker intended to refer to'—then she has effectively changed the subject; what we were interested in was the sort of reference that affects the truth or falsity of utterances. But if Mount is talking about reference *simpliciter*, then it is unclear what the 'on the speaker's part'-qualification could possibly mean. The truth-values of sentences or utterances aren't typically taken to vary between speakers and listeners; hence, what their sub-parts refer to shouldn't vary either. Besides, Mount's explicit notion of reference is an interpersonal one, given that it relies on what is maximally salient *both* the speaker and the listener. In short then, while Mount's response here seems intuitively plausible—at least on one interpretation—it also seems to be inconsistent with the rest of her theory. Her response therefore strikes me as rather puzzling.

perspective, etc. Each of these will plausibly influence what an individual finds maximally visually salient in a context, at least in certain circumstances. If this is right, however, then there is no such thing as visual salience relative to a context *simpliciter*. Rather, what there is is visual salience, in a context, relative to one or another perspective, or relative to a context-perspective pair.<sup>24</sup>

Similar considerations will apply to the rest of Gauker's criteria as well: being the most salient in prior discourse, being the most relevant, being the most clearly gestured at, being the most sensible in terms of fitting in a series, and being that which is most charitably understood as the speaker's intended referent are all plausibly properties that are sensitive *inter alia* to one's interests, background beliefs, and visual perspective. What this means for Gauker is that we aren't just faced with the task of weighing each of these criteria of salience against each other, so as to determine which object best fits them all; such weighing might often turn out to be messy, but a degree of messiness might nonetheless prove tolerable. Rather, the problem is that, in order to be able to weigh these six criteria against each other, we first need to find some way of combining different perspectives on a context into a single common perspective—so as to determine what, all things considered, is *the* most salient thing in the prior discourse, *the* object or property most clearly gestured at, etc., relative to the context and independent of any particular perspective on that context. I am skeptical that any such rubric for combining perspectives is in the offing. For one thing, it hardly seems plausible that perspectives on a context are the right sorts of things to be better or worse than each other, which might then underwrite their being weighted to a

---

<sup>24</sup>Note that this is equally an argument against Lewis (1979)'s thought that we can offer a notion of what is objectively salient in a discourse by appeal to the notion of a conversational scoreboard, which tracks the commitments that have been made over the course of the conversation thus far. The basic problem is that, while the conversational scoreboard may contain a number of beliefs, goals, questions under discussion, etc., it will never include a full perspective on the context. This means that the scoreboard will always, or at least almost always, leave under-determined what is maximally salient in context, since additional background beliefs, interests, and attentional states would be necessary to isolate one or another object as maximally salient. In other words, except perhaps in some very particular circumstances (such as contexts with only one object), there should always be multiple perspectives fully compatible with the conversational scoreboard, but which rank different objects as maximally salient relative to the context. If that's right, then a conversational scoreboard alone will not suffice to provide us with objective salience in context.

greater or lesser degree.<sup>25</sup> At best then, Gauker’s theory is incomplete. At worst, the theory cannot be completed. Given the depth of the challenges facing Gauker’s ‘best fit judgment’ theory, I suspect that the truth is probably the latter.

Finally we come to the third and most general objection to salientism: salientism cannot account for certain types of deceptive uses of language. Consider a case along the following lines. We are both employees at a large investment bank, and we have both just received an outrageously large bonus. Shortly after this, we bump into each other at work early in the morning, near a window that overlooks the nearly empty company parking lot. You query me regarding how I spent my bonus and I reply by pointing at the single small cluster of cars in the lot and uttering (2).

(2) I bought that.

This cluster is comprised of several rather ordinary cars and, in addition, a McLaren F1. I know that you are likely to take me to be talking about the McLaren, given just how much money we were just paid along with my well-known proclivity for purchasing flashy things. However, the car that I actually bought, and the one to which I intend to refer, is the Toyota Camry parked next to the McLaren. I am merely looking to have a bit of fun with you here.

It strikes me that my utterance of (2) is true in this instance, both in virtue my use of ‘that’ referring to the Camry and in virtue of my having actually purchased the Camry.<sup>26</sup> By any reasonable standard, however, it seems that the McLaren should be maximally salient here (in fact, a McLaren F1 is probably maximally salient more or less wherever it goes). Plausibly, the McLaren will be maximally salient to me, it will almost certainly be maximally salient to you, and by any reasonable standard it should be, objectively, the most salient

---

<sup>25</sup>Perhaps listener-perspectives that involve a failure to attend to the speaker’s utterance can be legitimately classified as ‘worse’ than other perspectives. Still, even discarding these will not be enough to ensure any sort of harmony among the plethora remaining perspectives—as was illustrated above with the woolly mammoth case above.

<sup>26</sup>One might, of course, doubt this judgment. Interestingly, rejecting it will turn out to entail some non-trivial negative consequences. I will introduce these shortly.

car in the lot.<sup>27</sup> So salientists of all stripes should be committed to the McLaren being the referent of my use of ‘that’. But if my utterance is indeed true, this cannot be the case. Rather, my use of ‘that’ must have referred to the Camry—as intentionalist theories look much better-positioned to predict.

Some may be inclined to dismiss this case as somehow non-standard, in virtue perhaps of its involving a (harmlessly) deceptive use of language. But note that similar cases can be generated without any hint of deceit by stipulating that the speaker, while sincere, is also a bit clueless. So, for instance, suppose that this cluster of automobiles actually isn’t all that far away and I therefore expect that my pointing gesture will be sufficiently clear to draw your attention to the Camry as opposed to the McLaren. Unfortunately, I have failed to properly account for the sheer visual magnetism of the McLaren F1. Thus, once more, your attention is naturally drawn to the McLaren—just as the attention of most other reasonable listeners in your shoes would be. And, in fact, I too am transfixed by the McLaren, even though I am neither pointing at it nor intending to refer to it. Once again, it seems that I have said something true of my Camry here, despite the fact that we are both attending to the McLaren, and despite the fact that, if anything is objectively salient in this context, it would seem to be the McLaren.

In fairness to salientists like Mount and Gauker, it should be noted that both of these judgments are, of course, dubitable. This might seem to put this objection on par with the salientist’s own argument from cases. But such reasoning would be mistaken. First of all, I suspect that judgments on the Camry-McLaren case are both clearer and more widespread than judgments on either Mount’s version of the Carnap-Agnew case or Gauker’s necktie case (particularly given how under-described the latter case turned out to be). Second,

---

<sup>27</sup>There is room to quibble about whether the McLaren is maximally salient to me, given that I intend to refer to my Camry. Perhaps an intention to refer to *o* entails that *o* is maximally salient to you. I think there is at least some reason to doubt this, however. Consider again the case of the rampaging mammoth. As noted in footnote 11, I can yell to you “Throw that spear at the mammoth!” while intending to refer to the spear I know to be next to you and while running full speed towards the hiding place I can see ahead of me. Plausibly, I take it, the hiding place in my visual field, not the spear that I know to be next to you, is the focus of my cognitive attention here. If that’s right, then we should reject this purported entailment.

the judgment regarding the sneaky version of the Camry-McLaren case can be reinforced by considering the following: regardless of whether the speaker is saying something true or false, she clearly isn't lying. The natural explanation for this would seem to be that the speaker hasn't said anything that she believes to be false. But if she is in fact referring to the McLaren, then she has indeed said something she believes to be false. What's more, she has said something about which she intends to deceive the listener. On any extant theory of lying, this is going to make it very difficult to predict that the speaker has not lied in this instance.<sup>28</sup> The problem, of course, is that the speaker plausibly hasn't lied here. Third, and finally, it turned out at least one version of intentionalism could in fact predict both Mount and Gauker's professed judgments regarding their respective cases. No such prospect would seem to be open to the salientist with respect to the Camry-McLaren cases.

This concludes my case against salientism. Contemporary salientist theories have tended to opt for either mutual maximum salience or objective salience as the relevant sort of salience for fixing reference in context. But mutual maximal salience ties utterance-truth and -falsity to the listener's mental state in an implausibly strong manner, whereas the thought that there is a single, objectively salient object relative to each context seems highly implausible given the degree to which salience can vary with background mental states, or what I have called a 'perspective'. Opting instead for listener-only salience views would only make the problem of listener-dependence worse, whereas speaker-oriented views entail, implausibly, that speakers must be occurrently attending to the referent at the precise moment of utterance.<sup>29</sup> Even if these various problems could be overcome, salience-based views of all stripes seem bound to make implausible predictions regarding both deceptive and slightly clueless speakers. Thus,

---

<sup>28</sup>Augustine (1952)'s venerable definition of lying was basically just this: a speaker lies to her listener iff she says that  $p$ , believes  $p$  to be false, and intends to deceive her listener with regard to  $p$ . Similar definitions have subsequently been proposed by Chisholm & Feehan (1977) and Williams (2002). More recently, a number of authors have come to doubt that this definition casts a sufficiently wide net; in particular, they have suggested that there are lies that are not intended to deceive (cf. Carson 2006, Sorensen 2007, Fallis 2009, 2012, and Saul 2012). Still, even these theorists agree that saying something that one believes to be false while intending to deceive one's listener with regard to the content of what one has said *suffices* to make an utterance a lie. What has been called into question is rather whether such deception is necessary.

<sup>29</sup>For examples illustrating the implausibility of this, see footnotes 11 and 27.

I urge that salientism of any sort be rejected.

## 2.7 Conclusion

I began by introducing the problem of reference fixing for demonstratives. While this is a sub-part of the wider project of associating uses of context-sensitive terms in general with semantic values—so as to generate predictions of utterance-truth and -falsity in context—it is often treated in isolation from that larger project. Adopting this restricted focus for the purposes of this chapter, I introduced the general outline of salientist theories, introduced the two best-developed versions of that view, and pointed to the range of other versions that are also available. I argued that the motivations for salientism, even purely within the realm of demonstratives, are far weaker than is standardly advertised. What's more, I showed that salientists of any stripe faces a number of serious, and possibly insurmountable, problems. On that basis, I concluded that salientism should be rejected. As the only serious alternative, this conclusion points towards the acceptance of some version of intentionalism.

In fact, I think that this conclusion should be welcomed. Among other things, intentionalism holds out the prospect of a more natural explanation of how a gap can arise between what a speaker's utterance means and what the listener, or even a semi-idealized listener, will take her to have meant. Since we have good evidence that these sorts of situations can in fact arise, we should refrain from endorsing a theory of meaning that entails that such situations should not be possible. Sometimes, it seems, speakers just aren't as communicatively conspicuous as they ought to be; indeed, sometimes they are even outright deceptive about what they are saying. But this should hardly be taken to mean that their utterances are true or false in ways contrary to what they believe. That requires mis-speaking. And, intuitively, mis-speaking is rather different from either misjudging what a listener will recover as the referent in context or intentionally speaking in such a way as to obscure what one is saying.

None of this, of course, should be taken to mean that there is any shortage of challenges

facing intentionalism. On the contrary, the challenges facing intentionalism are both many and serious. As we will see in the chapter 3, only a few versions of intentionalism are capable of handling cases of confusion like the Carnap-Agnew case. And, as we will see in chapter 4, even fewer versions can explain how a gap can arise between what the speaker intends to refer to and what a reasonable listener is likely to recover as the referent. Fewer still are capable of doing both. However, as I will aim to show over the course of those two chapters, there is at least one version of intentionalism—what I call the ‘constraint theory’—capable of meeting both these desiderata. What’s more, that version of intentionalism is able to account not just for the reference of demonstratives, but also for the reference of names, pronouns, and ‘true’ indexicals like ‘I’, ‘here’, and ‘now’.

# Chapter 3

## Reference and Confusion

*DROMIO OF SYRACUSE*      *What answer, sir? when spake I such a word?*  
*ANTIPHOLUS OF SYRACUSE*      *Even now, even here, not half an hour since.*  
*DROMIO OF SYRACUSE*      *I did not see you since you sent me hence,*  
*Home to the Centaur, with the gold you gave me.*  
*Shakespeare, The Comedy of Errors—circa 1590*

### 3.1 Introduction

In recent years in philosophy of language, there has been a renewal of interest in two topics: (i) intentionalism about the reference of singular terms like names and, in particular, demonstratives like ‘this’ and ‘that’<sup>1</sup>; and (ii) referential ‘confusion’.<sup>2</sup> What has somehow gone almost unnoticed in the literature is that these topics are deeply intertwined. For example, two classic hard cases in the theory of reference—e.g. Kaplan (1978a)’s ‘Carnap-Agnew’ case and Kripke (1977)’s ‘Smith-Jones’ case (more on these shortly)—are equally cases of referential confusion. The primary goal of this chapter will be to leverage this diagnosis

---

<sup>1</sup>Cf. Kaplan (1989a, 1990), Reimer (1991a, 1992), Bach (1992a,b), Siegel (2002), Fodor & Lepore (2004), Neale (2004), King (2013a,b), Heck (2013), and Speaks (2013).

<sup>2</sup>Cf. Millikan (2000), Camp (2002), Lawlor (2005, 2007), Scharp (2005), Schroeter (2007). Interest in confusion cases as such dates back further with scientific terms. See, in particular, Field (1973) and, more recently, Wilson (2006).

into a unified intentionalist theory of reference. The key to succeeding in that task will be to attend not just to what the speaker ultimately hopes to communicate, but also to how exactly she might go about trying to help the listener identify her intended referent.

To clarify my terms a bit: as was noted in chapter 1, by ‘intentionalism about reference’, I mean to refer to a class of theories unified by the claim that what determines the reference of expressions like names, demonstratives, and pronouns, relative to particular contexts, is some aspect of the speaker’s intentions in uttering the relevant term. Different intentionalist theories go different ways in terms of identifying just which aspects of the speaker’s intentions are relevant, and under what conditions these intentions succeed in fixing reference. We already touched on some of these differences in chapter 2; in the present chapter, we will explore these differences in far greater detail.

By ‘referential confusion’, on the other hand, I mean to refer to the sort of phenomenon that arises when agents conflate two different objects in some aspect of their mental lives. Crucially, this isn’t a matter of merely having some false beliefs about an object. Rather, referential confusion arises when an agent has multiple ways of thinking about an object that she takes to pick out one and the same object—but which in fact pick out different objects—and where she deploys these simultaneously in the course of either thinking a thought, forming an intention, or performing some other sort of mental act. Such circumstances make it difficult, if not impossible, to specify *which object* the relevant mental act was about.<sup>3</sup>

To illustrate the notion of referential confusion, consider a pair of cases. First, there is Kaplan (1978a)’s Carnap-Agnew case—a case that we already touched on in chapter 2. To review, the speaker (here, let us call him ‘Kaplan’) mistakenly believes that the picture hanging on the wall behind him is his picture of Carnap. Without turning to look, Kaplan points at the picture hanging behind him—a picture of Spiro Agnew—and says:

---

<sup>3</sup>For this way of characterizing cases of referential confusion, see in particular Camp (2002). While Camp and others have focused primarily on thinking thoughts about objects and engaging in inference on the basis of such thoughts, rather than on forming intentions, analogous considerations would seem to apply to any aspect of our mental lives that involve the engagement of our representational capacities. Generating object-directed intentions would seem to be just such an activity. For further discussion, see section 3.3.

- (1) That is a picture of one of the greatest philosophers of the twentieth century.

Here, Kaplan seems to have confused his picture of Carnap with the picture hanging on the wall behind him.<sup>4</sup> Nonetheless, the reference of Kaplan's use of the term 'that' seems clear enough: it refers to the picture of Agnew that is *actually* hanging behind him (Kaplan 1978, p. 239). Kaplan's confusion has thus led to a situation in which linguistic reference has parted ways from at least some aspect of his intentions.<sup>5</sup>

Similar phenomena can be observed with regard to names. Consider an example from Kripke (1977): Smith is in the distance raking leaves, but I mistake him for Jones. Modifying the case slightly so as to avoid any unnecessary and irrelevant entanglements regarding the semantics of question-answer pairs, suppose that now I point and say:

- (2) Jones sure is making a hash of his raking.<sup>6</sup>

Following Kripke, I take it that the speaker's (here, my) use of 'Jones' refers to some particular Jones in this context (Kripke 1977, pp. 263–64). The question is how we might explain reference to *that particular Jones* given that it is highly unclear that I intended to refer to any Jones at all, rather than to the man I saw raking leaves in the distance (i.e. Smith).

It is worth pausing to note just how different these cases turn out to be. In the latter, the name 'Jones' succeeds in referring to an absent object (i.e. Jones) despite the speaker's intention to refer to, among other things, the man raking leaves in the distance—a man who is actually present in the context of utterance. Kaplan's use of 'that', on the other hand, refers to the actual picture in his vicinity, in spite of what would seem to be Kaplan's fairly clear intention to refer to his (absent) picture of Carnap. The challenge then is to offer an

---

<sup>4</sup>Granted, this might seem to be a matter of false belief as opposed to confusion. As I explain in greater detail in subsection 3.3.1, however, this is only plausible at the level of belief, not at the level of intention.

<sup>5</sup>Some philosophers have taken the proper lesson of this case to be that we should simply give up on intentionalism. In fact, this was Kaplan (1978a)'s initial reaction, one which led him to endorse a theory where demonstrative reference was fixed by accompanying ostensive gestures. I offered arguments against both Kaplan (1978a)'s suggested demonstrationism and its salientist successors in the previous chapter.

<sup>6</sup>Kripke (1977)'s original case involves a brief colloquy: "What is Jones doing?" "Raking the leaves."

intentionalist theory of reference capable of accounting for both of these sorts of cases. As we will see below, no extant version of intentionalism is capable of meeting this challenge

This challenge is not insurmountable, however. In particular, my own ‘constraint theory’ is capable of meeting it. According to this theory, names and demonstratives bring with them different sorts of *public commitments*, which in turn serve to filter speakers’ diverse intentions for just those which satisfy these commitments. Thereby, the theory manages to sift through confused speaker’s myriad intentions in order to isolate a distinguished subset that serves to fix reference in context. Since names and demonstratives are associated with different public commitments, the theory obtains the flexibility necessary to deal with the sorts of cases just considered. What’s more, the constraint theory can be motivated by reflecting on the sorts of acts that uses of referential terms plausibly constitute. Essentially, cooperative speakers choose referential terms whose public commitments are satisfied by their intended referent in order to help *signal* to the listener that they intend to refer to that very object. While that signal alone typically won’t suffice to isolate a referent, it nonetheless provides a vital clue for the listener in her task of interpreting the speaker’s utterance.

Here is the plan for what remains: in section 3.2, I outline a very basic version of intentionalism about linguistic reference. In section 3.3, I return to the Carnap-Agnew and Smith-Jones cases in more detail. The goal will be to clarify both how these cases present a genuine challenge for intentionalists and to further clarify the relevant notion of confusion. Next, in section 3.4, I outline three ways in which intentionalists have tried to account for cases of confusion. I argue that, even restricting our attention purely to the domain of demonstrative reference, each of these ‘sophisticated’ versions of intentionalism exhibits serious flaws. In section 3.5, I argue that each of these proposals looks even worse if we try to extend it to account for names. If we aspire to offer a unified intentionalist account of referential terms—as I suggest we should—we must therefore look elsewhere. In section 3.6, I turn to the task of explicating the constraint theory. Not only is that theory capable of accounting for cases of confusion involving both names and demonstratives, it is also helps

to illuminate how speakers use referential terms in context to help convey their thoughts to listeners. Finally, in section 3.7, I conclude with a discussion of the important differences between how confusion manifests in thought and language, and what this might tell us about the nature of semantic content.

## 3.2 Simple Intentionalism

As was indicated above, intentionalism about reference is really a cluster of related theories, not a single account. Still, it will help to get a very basic version of intentionalism clearly in view, one that can serve as a common core for the various more sophisticated offshoots we'll consider below. 'Simple intentionalism', as I will call it, is the view that: (i) demonstratives refer when, and only when, they are used by speakers with a *referential intention*, or an intention that this term refer to some particular object; and (ii) whatever object figures into the content of this referential intention *just is* the referent of the relevant demonstrative, relative to the context of use.<sup>7</sup> The view can then be extended to account for the reference of names, pronouns, and whatever other expressions we might deem appropriate by appending those expressions to (i) and (ii) in the appropriate manner.

The problem now is to say more about what these referential intentions amount to. In subsection 3.4.3, in the course of discussing Reimer (1991a, 1992) and Bach (1992a,b), we will examine two serious attempts to make progress on the nature of referential intentions. For the purposes of defining simple intentionalism, however, let us assume that these intentions

---

<sup>7</sup>This view, at least as it applies to demonstratives, is often attributed to Kaplan (1989a) (cf. Bach 1992a, Siegel 2002, Wolter 2009). For reasons elaborated in footnote 15, this characterization of Kaplan's position probably isn't accurate. Rather, it is plausible that, at least as I have precisified the view here, no one has ever explicitly endorsed simple intentionalism in print. Numerous authors *have*, however, offered claims to the effect that 'the speaker's intentions/what she has in mind fix(es) the reference of X in context', without offering any further explanation regarding how this mechanism of reference fixing is supposed to operate. I think it not unreasonable to suspect that many such authors have had something like simple intentionalism in mind. For a representative sample, see: Donnellan (1966), Burge (1974), Kripke (1977), Bertolet (1980), Davies (1982), Lepore & Ludwig (2000), King (2001), Higginbotham (2002), Fodor & Lepore (2004), Cappelen & Lepore (2005), King & Stanley (2005), Szabó (2006), and Soames (2010). For some more sophisticated versions of intentionalism that won't receive adequate attention in this chapter, see Schiffer (1981), Neale (2004), Heck (2013), King (2013b), and Speaks (2013). I address some of these in chapter 4.

are just intentions to use particular terms to refer to particular objects. The content of these intentions might be overtly linguistic (e.g. “I intend with my use of ‘that’ that it refer to *o*”), or it might be less overtly linguistic (e.g. “I have *o* in mind and intend to refer to it—and to accomplish this I’m going to use the term ‘that’”).<sup>8</sup> Deciding between these proposals needn’t concern us here; for our purposes, all that matters is that the contents of these intentions are relations between uses of terms and objects (specified either *de dicto* or *de re*).<sup>9</sup> Simple intentionalism can thus be restated as follows: whenever a demonstrative or other context-sensitive referential term refers, it is accompanied by an intention the content of which is a relation between that use of the term and an object. That object, in turn, is predicted to be the referent.

Simple intentionalism, it should be noted, gets a number of things right. For instance, it allows that when I utter (1) while trying to point at the only picture in front of me, but where my aim goes slightly off, I still succeed in referring to that picture. Such cases are difficult to explain on earlier demonstration-based theories like those of Kaplan (1978a) and McGinn (1981). What’s more, as we saw in the last chapter, demonstration-based theories offer us little or nothing to say about the reference of uses of demonstratives that aren’t accompanied by the appropriate sort of physical demonstration. Simple intentionalism, in contrast, smoothly accounts for such cases.

---

<sup>8</sup>Note that if we try to make these less overtly linguistic—say, by making terms simply refer to ‘whatever the speaker has in mind’—we run into serious issues with sentences containing multiple referential terms. Would the speaker need to be occurrently thinking about each intended referent at the exact moment when she utters the appropriate term with which to refer to that object? Presumably not. But it is unclear how we might otherwise preserve the correlation of intentions and uses once we make these intentions non-linguistic. I have little doubt some other story is possible. Given that it has not yet been developed, however, I will assume throughout that simple intentionalists will avail themselves of either overtly or covertly linguistic intentions. Thankfully, nothing significant will hinge on this assumption below.

<sup>9</sup>Some may be tempted to think that intentions specified *de dicto* should not be allowed to fix reference. Recall, however, that our goal in this section is to develop the simplest, most minimally committal version of intentionalism we can. Ruling out intentions specified *de dicto* would thus cut against this spirit. What’s more, the Carnap-Agnew case already offers us good reason not to rule out the efficacy of intentions specified *de dicto*: the speaker’s intention to refer to the picture of Agnew must, it seems, be specified *de dicto* (as something like an intention to [refer to the only picture behind me]). Thus, if we hope to be intentionalists about reference even in cases like this one, we should be loathe to rule out the possible efficacy of intentions specified *de dicto*. These look to furnish our only real hope of explaining the Carnap-Agnew case!

Still, it shouldn't be too hard to see how confusion is going to pose a serious problem for simple intentionalism. In cases of confusion, we find ourselves with too many things competing to fill the object-place in the content of the speaker's referential intention. That fact is going to leave the simple intentionalist either saying something false about how reference falls out in these cases or else saying nothing at all about them. To get this problem better in view, let us return to the cases of confusion introduced at the start of this chapter.

### 3.3 Two Cases of Speaker Confusion

In this section, I re-introduce the Carnap-Agnew and Smith-Jones cases in slightly more detail, elaborating on why each constitutes a case of referential confusion and how each poses a serious challenge for simple intentionalism.

#### 3.3.1 The Carnap-Agnew Case

Here is how Kaplan (1978) originally introduces the Carnap-Agnew case:

Suppose that without turning and looking I point to the place on my wall which has long been occupied by a picture of Rulolf Carnap and I say:

- (1) That is a picture of one of the greatest philosophers of the twentieth century.<sup>10</sup>

But unbeknownst to me, someone has replaced my picture of Carnap with one of Spiro Agnew...I have said of a picture of Spiro Agnew that it pictures one of the greatest philosophers of the twentieth century. (Kaplan 1978a, p. 239)

Following Kaplan, I take it that the speaker in this case (i.e. Kaplan) has uttered a falsehood. More specifically, Kaplan's utterance is false in virtue of the fact that his use of 'that' refers to the picture of Angew in this context—meaning that (1) serves to predicate something false of that picture (assuming, of course, that Agnew wasn't writing philosophy on the sly, perhaps under the pseudonym 'W.V.O. Quine' or 'Richard Montague').

---

<sup>10</sup>I have once more taken the liberty of replacing Kaplan (1978a)'s 'dthat'-operator with the ordinary English 'that'.

Unsurprisingly, not everyone has agreed with Kaplan’s judgment on this case. In particular, King (2013a) has recently claimed that this is in fact a case of reference failure, which in turn explains our judgment that the utterance is false. I will offer arguments against that claim in subsection 3.4.2. For the time being however, I am only interested in why this case proves problematic for the simple intentionalist. To that end, it won’t much matter whether this is a case of successful reference to, and false-saying about, the picture of Angew, or whether it is a case of reference failure.

Here is how the case proves problematic for the simple intentionalist: plausibly, Kaplan intends to refer to his picture of Carnap. After all, he was presumably *trying* to say something true. But Kaplan also plausibly intends to refer to the picture hanging behind him, i.e. the picture of Agnew.<sup>11</sup> The problem for the simple intentionalist is to explain why Kaplan’s utterance is false. To do that, she would need to predict either that reference fails here, or else that Kaplan’s use of ‘that’ refers to just the picture of Agnew. Unfortunately, the resources that simple intentionalism has to offer aren’t up to making either of these predictions.

Consider what the simple intentionalist will have to say about this case: according to (i), demonstratives refer when and only when they are used by speakers with the intention that they refer to some object. Since Kaplan has at least one such intention in this case, (i) entails that his use of ‘that’ must refer to something. The question is: what does it refer to? According to (ii), a particular use of a demonstrative refers to whichever object figures into the content of the referential intention accompanying it—that is, the intention which satisfies (i). The problem in this case is that Kaplan has not just one intention to refer, but two distinct intentions—each pointing to a different object. Since (ii) presupposes that the speaker has only one referential intention in using a demonstrative, it is unclear what a simple intentionalist should predict when there is more than one such intention. In the

---

<sup>11</sup>In case one is inclined to think that only intentions specified *de re* matter for fixing reference: perhaps Kaplan in fact *saw* the picture of Agnew on entering his office. He simply failed to notice the switch. In this case, he would seem to be capable of having an intention to refer to the picture of Agnew, specified *de re*. He simply conflates this with his intention to refer to his picture of Carnap, also specified *de re*.

Carnap-Agnew case, (i) forces her to predict that Kaplan’s use of ‘that’ refers to *something*. The choices, then, would seem to be either that Kaplan’s use of ‘that’ refers to both pictures, or else that which picture it refers to is indeterminate. Picking either picture at the expense of the other would be intolerably *ad hoc*. But if Kaplan’s use of ‘that’ refers to both pictures, then his utterance would seem to be, in certain respects at least, both true and false—not just false. And if reference is indeterminate here, then the truth-value of Kaplan’s utterance would presumably be indeterminate as well. Neither of these look like good predictions.

One final thought on the case: some may be tempted to say that Kaplan was only thinking about his picture of Carnap, and thus that this is a case of mere false belief rather than a case of referential confusion. But even supposing that this were true, this clarity in thought is by no means guaranteed to translate into clarity in intentions.<sup>12</sup> That is, regardless of whether Kaplan is genuinely confused at the level of thought or whether he merely has some false beliefs about the world, it seems perfectly appropriate to characterize him as both intending to refer to the picture of Carnap and intending to refer to the picture of Agnew. Thus, even granting that certain versions of the Carnap-Agnew case plausibly fail to involve confusion at the level of thought, such cases will still involve genuine confusion at the level of intention. For our purposes, all that we are really interested in is confusion at the level of intention, since it is confusion at this level that poses the most direct challenge to intentionalism.

### 3.3.2 The Smith-Jones Case

Turning now to the Smith-Jones case, recall that the speaker sees Smith in the distance raking leaves and mistakes him for a particular Jones. She thus points and says:

(2) Jones sure is making a hash of his raking.

---

<sup>12</sup>See footnote 11 for reasons to doubt the antecedent here—at least on certain versions of the case. If Kaplan saw the picture of Angew on his way in but failed to notice the switch, he will plausibly be thinking about both pictures *de re*. That is, he will be confusing them for each other, not just falsely attributing a property to one of them. What this indicates, I take it, is that there are indeed versions of the Carnap-Agnew case that involve genuine confusion at the level of thought, not just confusion at the level of intention.

The claim was that the speaker’s use of ‘Jones’ succeeds in referring to the Jones whom she mistakenly believes to be raking the leaves in the distance, *not* to the man who is actually raking the leaves (i.e. Smith). Thus, (2) is true iff that particular Jones is making a hash of his raking, wherever he is; the truth or falsehood of (2) has nothing at all to do with Smith.

One point of clarification: the sort of reference in which we are interested here is what we might, following Kripke (1977), call ‘semantic reference’. With Kripke, I take it that:

“Jones,” in the common language of both [the speaker and listener], is a name of Jones; it never names Smith. (Kripke 1977, p. 263)

In other words, for a use of ‘Jones’ to semantically refer to someone, that person must actually be named ‘Jones’. This restriction, however, should not be understood as preventing the speaker from using the term ‘Jones’ to *speaker-refer* to Smith, so long as all that we mean by this is that there is some sense in which the speaker *intended* to refer to Smith with her use of the name ‘Jones’. The point is that the presence of such an intention would seem to have no effect the truth or falsity of the utterance itself. Crucially, our inquiry here is directed at the sort of reference that helps to explain the truth or falsity of sentences in contexts.<sup>13</sup> That is, at least on the standard understanding of these notions, what we are interested in is *semantic reference* rather than *speaker reference*.

The Smith-Jones case threatens simple intentionalism in a way that mirrors the Carnap-Agnew case above: the speaker’s use of the name ‘Jones’ clearly refers to Jones, but nonetheless the speaker plausibly intends to refer *both* to Jones *and* to Smith. That is, she has conflated the two—so that, in intending to refer to the man in the distance raking leaves (i.e. Smith), the speaker also intends to refer to Jones and vice-versa. The problem, once more, is that simple intentionalism fails to offer the resources necessary for teasing these intentions apart, so as to determine which of them fixes reference in this context. In other words, given that the speaker seems equally intent on referring to Smith and referring to

---

<sup>13</sup>Recall that, as noted in chapter 1, footnote 2, I am using ‘utterance’ and ‘sentence in context’ interchangeably, as whatever meaning and truth-conditions properly attach to in context.

Jones, we once more lack an explanation for why, semantically at least, she succeeds only in referring to Jones.<sup>14</sup>

Having shown that simple intentionalism cannot adequately account for cases of referential confusion, I turn now to the question of whether any of the more sophisticated versions of intentionalism that have been developed fare any better in this regard.

## 3.4 Sophisticated Intentionalisms

In this section, I explore three different strategies for modifying intentionalism so as to handle cases of referential confusion. I argue that, even restricting our attention only to cases involving demonstratives, each of these theories ultimately looks unappealing.

### 3.4.1 The Occurrent Perception Theory

Kaplan (1989a) and Siegel (2002) both suggest that the scope of simple intentionalism should be restricted to just ‘perceptual uses’ of demonstratives (Kaplan 1989a, p. 582; Siegel 2002, p. 1).<sup>15</sup> To count as a perceptual use of a demonstrative, the speaker must not only be

---

<sup>14</sup>Some readers may be tempted to follow Kaplan (1990) in thinking that names are individuated by their referent, essentially making ‘generic’ names like ‘Jones’ highly ambiguous. This would allow for the following type of response to this case: only one of the speaker’s intentions here is properly an intention to disambiguate the name ‘Jones’ in accord with the objects to which it can potentially refer. But consider that when a speaker intends to use the name ‘Jones’ just to refer to Smith (being under the false impression, perhaps, that this is Smith’s name—not mistaking him for a particular Jones), reference fails. Thus, it is not enough to say that the speaker only has one relevant disambiguating intention; she has one relevant disambiguating intention that would lead to referential success, and another that would lead to reference failure. We still require an explanation for why it is the former that figures in the semantics of the Smith-Jones case. For more extensive discussion of Kaplan’s theory of names, see chapter 4.

<sup>15</sup>Kaplan (1989a) is often characterized as advocating a simple intentionalist picture as opposed to a more restricted theory like this one (cf. Bach 1992a, Siegel 2002, Wolter 2009; for a counter-instance, see Reimer 1991a). Here is what Kaplan actually has to say on the subject:

I am now inclined to regard the directing intention, *at least in the case of perceptual demonstratives*, as criterial, and to regard the demonstration as a mere externalization of this inner intention. (Kaplan 1989a, p. 582; italics mine)

Although it is possible to interpret Kaplan as being open to the possibility that simple intentionalism might be correct *simpliciter*, he is careful not to commit himself to that view. In light of this rather overt caveat, it strikes me as highly uncharitable to attribute to Kaplan an unalloyed simple intentionalist view.

occurrently perceiving her intended referent—that is, she must not only be perceiving the relevant object at the moment of her utterance—her intention to refer to that object must also be grounded, at least in part, in this perceptual connection. Thus, according to Kaplan and Siegel, demonstrative reference isn't fixed by just any old intention to refer; rather, insofar as reference is fixed by intentions at all, it is fixed by *perceptually grounded* ones.

This 'occurrent perception theory' offers several advantages. First, it can deal with a large class of confusion cases as follows. Suppose that the speaker intends to refer to both  $o_1$  and  $o_2$ . So long as she is only perceptually discriminating  $o_1$ , not  $o_2$ , at present, then it doesn't matter if the speaker is conflating  $o_1$  and  $o_2$ . Only her intention to refer to  $o_1$  will count as perceptually grounded, and thus the speaker will only succeed in referring to  $o_1$ . Second, the theory can effectively sidestep the Carnap-Agnew case, since that case does not involve any occurrent perception of either of the intended referents. The case, quite simply, falls outside of the domain of the theory. Thus, it cannot serve as counterexample to the present theory.<sup>16</sup>

This second advantage is also simultaneously the primary *disadvantage* to this approach. The occurrent perception theory fails to account for large swaths of our ordinary use of terms like demonstratives. We should ask, therefore, whether there isn't some more general theory that can match the occurrent perception theory in terms of the accuracy of the results it can deliver. Assuming that we can get such a theory on the table, that theory would, in virtue of this greater generality, seem to be preferable to the occurrent perception theory. The occurrent perception theory might even constitute a special case of that more complete theory. As it happens, we won't have to look far to find a superior theory of reference.<sup>17</sup>

---

<sup>16</sup>Kaplan (p.c.) has suggested that the occurrent perception theorist might also allow that 'just perceived' objects can guide speaker's referential intentions as well—thus effectively extending the theory to account for certain versions of the Carnap-Agnew case. The basic idea is that recent perceptual memory is similar enough to occurrent perception to ground reference. Unfortunately, this extension only goes so far: if Kaplan happened not to have seen the picture of Agnew on the way into his office, or if he has simply been sitting around at his desk long enough, the view once more offers no verdict on the case. Since this modification introduces significant complexity (just what are the relevant bounds of memory for the purposes of reference fixing?) for only a minimal payoff, I leave it to the side in the main text.

<sup>17</sup>The occurrent perception theory faces another serious problem as well: plausibly, it faces counterexam-

### 3.4.2 The Coordination Account

King (2013a) has recently adopted a different strategy for modifying simple intentionalism:

I suggest we say that the value of a use of a demonstrative in a context is that object *o* that meets the following two conditions: 1) the speaker intends *o* to be the value; and 2) a competent, attentive, reasonable hearer would take *o* to be the object that the speaker intends to be the value. We can abbreviate this by saying that an object *o* is the value of an occurrence of a demonstrative in context just in case the speaker intends *o* to be the value and the speaker successfully reveals her intention. I'll call this account of how demonstratives acquire values the coordination account. (King 2013a, p. 13; underlining in original)<sup>18</sup>

King's basic suggestion is that successful reference requires not just an intention to refer to some object, but also the determination that a suitably idealized listener would coordinate with the speaker on that object as the referent. In other words, King seeks to constrain simple intentionalism by appealing to the *communicative potential* of particular speech acts.

As it stands, the coordination account alone isn't enough to help us predict reference in cases of confusion. The problem is that, in such cases, the speaker will have multiple candidate *o*'s to which she intends to refer, one or more of which might be recoverable by the listener in context. Being well aware of such cases, King (2013a) offers the following addendum to his theory: whenever speakers simultaneously intend to refer to multiple, distinct objects, reference fails (p. 16). Amended in this fashion, there is much to like about King's coordination account. It can deal with the Carnap-Agnew case by noting that, since the speaker has conflicting intentions, reference fails. That prediction, combined with the assumption that, in sentences involving neither implicit nor explicit negation, reference

---

ples even within the highly restricted domain for which it purports to account. Consider a case in which Kaplan is facing the picture of Carnap, which this time is hanging *next to* a picture of Agnew. Due to some temporary fluke in his sensory-motor system, however, Kaplan's aim is off and when he utters (1) he points directly at the picture of Agnew (perhaps he is so close as to actually be touching that picture). This is despite his perceptually-grounded intention to refer to the picture of Carnap. It seems to me clear that, *pace* what the occurrent perception theory predicts, Kaplan's use of 'that' does not refer to the picture of Carnap here. Whether or not it successfully refers to the picture of Agnew is a different matter.

<sup>18</sup>Note that King uses the term 'value' here in the same sense that I have been using the term 'referent' (cf. King 2013a, p. 2, footnote 2). See chapter 2, footnote 4 for more on this usage.

failure entails falsehood, suffices to explain why Kaplan’s utterance of (1) is false. What’s more, King’s theory is both broad in its scope and grounded in a simple and appealing idea: the theoretical role of semantic content is to help explain communicative potential.

Unfortunately, King’s coordination account suffers from several some serious drawbacks. Here, I focus on just two.<sup>19</sup> First, *pace* King, there is reason to think there is no such thing as ‘what a competent, attentive, reasonable listener (who knows the common ground, King 2013a adds at pp. 14–15) would recover as the referent in context’, at least in most contexts. Essentially, this worry parallels the one I raised for Gauker (2008)’s objective salience theory in chapter 2. For Gauker, the problem was that salience depends on perspective in a way that significantly complicated the possibility of aggregating individual perspectives into a single common perspective. Similarly, even idealized listeners will require a perspective on the context: that is, they will need to be assigned a set of background beliefs, interests, a visual perspective, etc. Only beings with properties like these will be capable of going about the task of interpretation—i.e. reference-recovery—in anything like the way that ordinary listeners do. But once we grant idealized listeners perspectives, there is reason to doubt that there will be general agreement among such listeners regarding which object should be taken to be the referent in context; more plausibly, what an idealized listener is going to recover is going to depend on her particular perspective on the relevant context.

To better illustrate the problem here, it will help to consider a specific case. Suppose that we are standing together in a park, looking out over a large field in which a number of dogs are playing frisbee. Gesturing only vaguely in their direction, I say:

(3) That dog is excellent at catching frisbees.<sup>20</sup>

---

<sup>19</sup>For discussion of another worry, see Speaks (2013). Speaks objects to King’s theory on the grounds that, since King hasn’t specified how to interpret the modal ‘would’ in his clause 2, the theory is incomplete as it stands. Speaks goes on to argue that, on any of the standard ways of evaluating the modal—and even some fairly non-standard ones—King’s theory delivers counterintuitive results in a wide range of cases. King disputes this in his (2013b). For the purposes of this chapter, I leave this worry to the side.

<sup>20</sup>King discusses an analogous case, involving a crowd of people on the Venice pier, in his (2013a). This sort of case was originally introduced in Reimer (1991a,b) and involved dogs and frisbees.

King claims that this sort of case involves reference failure, since even a suitably idealized listener won't be able to determine *which* dog the speaker intended as the referent (King 2013a, p. 8). With so little detail regarding the case, this prediction may well sound reasonable.

Suppose, however, that one of the dogs (call her 'Champ') is noticeably better than the rest at catching frisbees. Suppose further that the speaker intends to refer to Champ. King's earlier claim now seems intuitively wrong; a suitably idealized listener presumably *would* recover Champ as the referent in this version of the case. Not only that, it also seems right to say that the speaker's utterance of (3) was true in these circumstances—meaning that reference must have succeeded. Given the circumstances, the speaker's use of 'that dog' presumably succeeded in referring to Champ.

Still, it is far from clear that *every* suitably idealized listener would recover Champ as the referent in this scenario. Let us spell out the case in yet more detail: suppose that the common ground contains nothing about the listener's initial attentional state.<sup>21</sup> Now suppose that the idealized listener is attending to nothing in particular at the time of the utterance (other than to the speaker's utterance itself). In that case, this idealized listener will plausibly identify Champ as the referent. But suppose instead that our idealized listener happens to be attending to a different dog ('Rover') at the time of the utterance. Suppose further that Rover is catching frisbee after frisbee *pretty darn well*, though not quite as well as Champ. If the listener is sufficiently engrossed watching Rover, and if Rover is sufficiently good at catching frisbees, the listener—no matter how idealized—may well fail to notice Champ. In that case, this still-idealized listener is likely to recover Rover as the referent.<sup>22</sup>

---

<sup>21</sup>I take it that speakers often utter sentences like (3) partly in order to prompt their listeners to attend to particular objects, not because they already know what those listeners are attending to.

<sup>22</sup>At this point, King might claim that we have deviated from playing by his rules; an idealized listener, after all, should be capable of attending to *everything* at the context. So the listener I have concocted will not count as having been suitably idealized. But notice that no human being, no matter how idealized, is capable of attending simultaneously to everything at a context—certainly not simultaneously, and probably not ever. So King's hypothetical demand would serve to push his idealized listeners past the bounds of idealized *human* listeners. I take it that, if we want to learn about the sorts of natural languages that humans speak, we would probably be well-advised to constrain any idealizations we invoke in semantic theory to idealizations within the realm of idealized human beings—not idealized beings of a sort whose qualities go beyond what at least a few among us might actually hope to achieve through concerted practice and effort.

What this case illustrates is that we have reason to doubt that there will ever be any one thing that a suitably idealized listener would recover, in context, as the referent—at least in contexts that contain more than one object. What a suitably idealized listener will recover in context plausibly depends on her perspective, but perspectives aren't the right sorts of thing to be more or less ideal (at least once we specify that the listener is in fact paying sufficient attention to the speaker). If there is no one ideal perspective, however, then there will likewise be no one thing that any idealized listener will recover as the referent.<sup>23</sup> King's proposal thus faces a serious problem: since the coordination account requires that, for reference to succeed, there must be *a single* object that a suitably idealized listener would recover as the referent, and since almost never will there be such an object, the coordination account looks bound to over-generate predictions of reference failure. For King, reference fails not just in the Carnap-Agnew case, for instance, but also in the case where the speaker intends to refer to Champ and where Champ is clearly the best frisbee-catcher at the park.<sup>24</sup>

The second worry is simpler: even granting King everything he wants with respect to idealized listeners, there is reason to think that the coordination account gets a number of

---

<sup>23</sup>King might try weakening his condition 2 so as to make it a claim about what *most* suitably idealized listeners would recover in context. But given the myriad different perspectives that we can imagine attributing to a suitably idealized listener, I think we should be skeptical that such a weakened version of the theory would remain tractable. The problem, in short, becomes one of trying to evaluate 'most' in an infinite model.

<sup>24</sup>In his (2013b), King adds to his condition 2 that the idealized listener should have no mental states relevant to how she will recover the referent in context that aren't specified in the common ground (p. 14, footnote 9). He adopts this new constraint in response to a case that he attributes to David Manley: imagine a listener with an intentionometer, who can accurately recover the speaker's referential intentions whatever she actually says (Manley's version of the case involves a hyper-attentive CIA agent, but the details are unimportant here). The first thing to note about this is that King's new constraint fails to rule out some fairly bizarre cases of reference, so long as it is common knowledge that the listener possesses this intentionometer. Consider, for example, the case from chapter 1 where I point to a table and say 'that' while intending to refer to Normore. On King's theory, reference succeeds here. This strikes me as an odd prediction. Second, and more importantly for our purposes, King now seems to be positing that his idealized listeners have no perspective on the context beyond what is assigned to them by the common ground. But note that this may amount to little, or even nothing at all. In the Champ-Rover case, for instance, nothing about the listener's initial attentional state was specified in the common ground—and yet *no* initial attentional state turned out to be neutral with regard to what the listener would recover as the referent. If the listener is initially focused on most anything aside from Rover, she will recover Champ as the referent. If she happens to be focused on Rover, then she will recover Rover. Without going beyond the common ground to assign the idealized listener a perspective on the world, we simply cannot say what she would recover as the referent in this, or in most other, contexts.

cases wrong. Consider a variant of the Carnap-Agnew case in which Kaplan mistakenly believes that Carnap made a brief and extraordinarily unsuccessful foray into politics. Once more, Kaplan points behind himself at what he takes to be his picture of Carnap (but which is in fact a picture of Agnew). This time, however, he utters (1') instead of (1).

(1') That is a picture of one of the worst politicians of the twentieth century.

Plausibly, Kaplan's utterance is true in this instance—not true in the way he thinks, but true nonetheless.<sup>25</sup> But the coordination account predicts otherwise; it predicts that this is a case of reference failure. And given the absence of any sort of negation in (1'), this prediction means that the coordination account is stuck predicting that Kaplan has uttered a falsehood here. This, I take it, is a serious strike against the coordination account.<sup>26</sup>

We have thus seen two reasons to reject the coordination account: first, it is highly unclear that conflicting intentions always lead to reference failure. Even if they do, the coordination account still risks predicting reference failure in a wide swath of cases where it shouldn't—since even idealized listeners can diverge with respect to what they recover as the referent in context, depending on their perspective on the world

### 3.4.3 Structured Intentionalism

Reimer (1991a, 1992) and Bach (1992a,b) introduce a pair of closely related views aimed at dealing with confusion cases by examining what we might think of as the *structure* of speakers' intentions in more detail. Reimer proposes a distinction between what she calls

---

<sup>25</sup>There are numerous other variants on this case. For instance, the setup could be as in the original Carnap-Agnew case with the exception of the picture of Carnap being replaced not with a picture of Agnew, but rather with one of Ruth Barcan Marcus. In that case, Kaplan's utterance would once again seem to be true, though only inadvertently so. In language, as elsewhere, we should be loathe to denigrate lucky success.

<sup>26</sup>King (p.c.) has contesting the judgment that Kaplan's utterance is true in this case. As further evidence, King has claimed that we would hesitate to say of Kaplan that "He said of the picture of Agnew that it is a picture of one of the worst politicians of the twentieth century." I can only report that I feel no such compunctions. Interestingly, even King himself takes judgments like these—to the effect that Kaplan's utterance is in fact true—seriously enough that he has recently offered an alternative version of his account that is specifically designed to vindicate them (cf. King 2013b).

*primary* and *secondary* intentions, whereas Bach introduces a distinction between a speaker's *overall communicative intention* and the properly *referential* sub-part of it. While these strategies may sound rather different, the core idea is in fact quite similar. I will start by examining Reimer's view in more detail. Then I will turn to Bach's account. Finally, I will explain why both these views are, in a crucial sense, incomplete.

Reimer (1991a, 1992) proposes a theory, which she calls 'quasi-intentionalism', according to which reference is fixed by the speaker's intention to *demonstrate* an object. Crucially, Reimer's notion of 'demonstrating' is not equivalent to merely pointing or gesturing at. Rather, what Reimer has in mind is that, in using demonstratives, speakers typically aim to present a particular object to the listener *as* the referent. Pointing, for Reimer, is merely one among the many tools we use to help demonstrate objects to each other. In other circumstances, we might instead rely on an object's antecedent salience to single it out as our intended referent. This, in conjunction with the choice of 'this' or 'that' in context, would still count for Reimer as 'demonstrating' the object (Reimer 1991a, p. 199–200).<sup>27</sup>

In her (1992), Reimer further distinguishes between the speaker's *primary* intention—i.e. her intention to refer to some object that initially motivates her to say something to the listener—and her *secondary* intention—i.e. her intention to demonstrate this object in a particular way (pp. 389–90). Reimer's thesis can thus be rephrased as: it is the speaker's *secondary* intention that serves to fix demonstrative reference.

Here is how Reimer applies this distinction to the Carnap-Agnew case. There, she claims that what Kaplan's use of 'that' refers to is determined by:

[his] 'secondary' intention: an intention to demonstrate, and say something of, the picture in the general direction of the gesture—a picture which (unbeknownst to Kaplan) is not the picture of Carnap. (Reimer 1992, p. 389).

---

<sup>27</sup>Strangely, Reimer is often grouped with those who appeal to either salience or gestures in order to fix the reference of demonstratives (cf. Siegel 2002, King 2013a). This probably derives from Bach (1992a)'s influential criticism of an earlier paper, namely Reimer (1991b), where Reimer does seem sympathetic to some sort of salience-based theory. However, by the time of Reimer (1991a, 1992), Reimer had clearly come around to endorsing a version of intentionalism. To be fair, Reimer's quasi-intentionalism does obliquely appeal to salience as one factors that speakers sometimes take into account in formulating their intentions; still, this hardly puts the view on par with the true salientist views we examined in chapter 2.

The view thus entails that when speakers intentionally gesture in the course of using demonstratives, those gestures will figure prominently in their secondary intentions—that is, in their intentions to demonstrate particular objects. They must therefore commit to some particular way of using that gesture to demonstrate the relevant object. For instance, the speaker might commit to referring to the such-and-such in the general direction of her gesture. In the Carnap-Agnew case, the picture of Agnew, but not the picture of Carnap, is just such a such-and-such (here, picture) in the direction of the speaker’s gesture.

A potentially serious problem arises at this point: suppose that we follow Reimer in thinking that whenever speakers point, they do so with the intention to demonstrate an actual such-and-such the direction of that pointing. Then the theory will be unable to account for cases of ‘deferred reference’—or, roughly, cases which involve present objects somehow standing in for absent ones.<sup>28</sup> Here is an example from King (2013a):

Suppose we are in a graduate seminar. It is common ground that a participant, Glenn, is skiing at Mammoth today and so couldn’t attend the seminar. It is also common ground that he always sits in a now empty chair to my right. Intending to say something about Glenn, I point at the empty chair and say “I bet he is having fun right now.” (King 2013a, p. 15)

Setting aside the fact that a gendered pronoun has been used here rather than a demonstrative *per se*, we can easily see how this sort of case threatens Reimer’s proposal: the problem is that Glenn isn’t in the direction of the speaker’s pointing gesture. Yet King’s use of ‘he’ plausibly succeeds in referring to Glenn nonetheless. Intuitively, the problem is that, as it stands, Reimer’s proposal is *too restrictive* when it comes to how we can use gestures to direct listeners to demonstrata. The question is how one might loosen this aspect of Reimer’s theory while still preserving the essential features of her view. One option would be to allow that speakers are in fact capable of demonstrating objects by pointing with slightly different intentions than those alluded to above: intentions to demonstrate the such-and-such that

---

<sup>28</sup>Discussion of deferred reference dates back at least to Quine (1968). For further discussion, in addition to an impressive catalogue of cases, see Nunberg (1995).

stands in one or another representational relation  $R$  to a such-and-such in the direction of an ostensive gesture. This, I take it, is probably the more generous way of understanding Reimer's theory. An interesting side-effect of this interpretation is that it makes Reimer's theory coincide almost exactly with Bach (1992a,b)'s neo-Gricean proposal.

According to Bach (1992a,b), what fixes reference relative to a context are the speaker's specifically *referential intentions*. Here is Bach on the features exhibited by such intentions:

The relevant intention here, the specifically referential one, is part of a communicative intention, a reflexive intention whose distinctive feature is that 'its fulfillment consists in its recognition' by one's audience, partly by supposing (in Gricean fashion) that the speaker intends his intention to be recognized...[A] referential intention isn't just any intention to refer to something one has in mind but is the intention that one's audience identify, and take themselves to be intended to identify, a certain item as the referent by means of thinking of it a certain identifiable way. (Bach 1992a, p. 144)<sup>29</sup>

Referential intentions, for Bach, are thus intentions to make a particular object *evident* as the referent. Part of what having such an intention requires is conceiving of a particular way for the audience to identify that referent. In addition, the speaker must intend to make this very intention evident to the listener—so that the listener can see that he is supposed to identify the referent in this way on the basis of the speaker's intention that he do so.<sup>30</sup>

Crucially for Bach, pointing is just one of the ways in which we signal our intentions to each other; there is nothing particularly special about it (Ibid., p. 144). Thus, as on our considered understanding of Reimer's theory, speakers needn't intend for their listeners to identify the referent directly, by means of a pointing gesture. Rather, the speaker may intend for the listener to rely on other sorts of cues in order to identify the referent, or to

---

<sup>29</sup>Confusingly, Bach (1992a) sometimes talks of reference 'succeeding' only when the listener actually identifies the referent that the speaker's referential intention serves to isolate (Bach 1992a, p. 144). I take it that the notion of success that Bach has in mind here isn't the one that we are interested in—one that helps to determine reference failure, and thus truth or falsity in context—but is rather a more social notion—something like 'playing the desired role in a communicative exchange'. Note that if Bach means 'success' here in a sense that is relevant to utterance-truth or -falsity, then this would entail, implausibly, that, whenever a listener fails to pay adequate attention, the speaker's utterances will shift in its truth-values. This would mean that, *pace* his many claims to the contrary, Bach's theory isn't a purely intentionalist theory after all.

<sup>30</sup>In adopting these requirements, Bach (1992a) is essentially adapting Grice (1989d)'s theory of occasion meaning to the particular case of reference. For more extensive discussion, see chapter 4.

recognize that the intended referent is only loosely related (again, via some representational relation  $R$ ) to the object the speaker actually pointed to. Thus, Bach is able to handle the Carnap-Agnew case by claiming that, while Kaplan may in some sense intend to refer both to his picture of Carnap and to the picture behind him, only the latter constitutes a genuinely referential intention. This follows from the fact that Kaplan can only reasonably expect for the listener to identify his intended referent in these circumstances by following his pointing gesture and identifying the picture behind him (i.e. Kaplan) as the referent (Bach 1992b, pp. 296–97).<sup>31</sup> Likewise, Bach can account for King’s empty chair case by claiming that the speaker intends for the listener to identify Glenn as the referent by reasoning about her utterance, her gesture, and the shared belief that Glenn is off skiing today.

For all of the virtues of these proposals, there are two serious problems facing each.<sup>32</sup> First, both Reimer and Bach’s proposals implicitly rely on the speaker having only *reasonable* beliefs about the context. Consider what happens to both their explanations of the Carnap-Agnew case if we stipulate that Kaplan falsely believes that the listener is a mind-reader. Then, according to the Reimer’s theory, Kaplan might have a genuine secondary intention to demonstrate his picture of Carnap—since he will believe that, merely in virtue of his forming this intention, the listener will be in a position to identify his intended referent. But then he should succeed in referring to that picture in these circumstances! Something similar holds for Bach’s theory, since if Kaplan believes that his listener is a mind-reader, he *can* form a

---

<sup>31</sup>Bach offers a similar take on this case in Bach (1987, pp. 182–86).

<sup>32</sup>Despite treating them in parallel in the main text, there are actually a few cases where Bach (1992a,b) and Reimer (1991a, 1992)’s views come apart, predictively speaking. For instance, Bach faces the following sort of problem—while Reimer does not. Suppose that I am reading the newspaper, which I know that you haven’t yet seen. I exclaim “My word, you have to see this!” knowing full well that you have not, and thus that *there is no way* in which you might be able to identify my intended referent, at least at present. Still, my use of ‘this’ plausibly succeeds in referring to the story that I have in mind. Bach might try to account for this case by allowing that the relevant sort of referential intentions can extend over time, and can include intentions to supplement one’s original utterance with further linguistic or non-linguistic material (cf. Clark & Wilkes-Gibbs 1992). Alternatively, Bach might appeal to a looser version of his notion of a ‘way of identifying the referent’, such as: whatever the speaker intends to refer to with this use of a demonstrative. While this latter option might initially seem simpler, it also risks entailing that we should be able to use demonstratives to refer to whatever we like on any occasion—since listeners will presumably always have recourse to at least this minimal way of identifying the referent. Whichever way one goes, Bach’s theory, in contrast to Reimer’s, is going to have to be further fleshed-out if it is going to handle this sort of case.

genuinely referential intention to refer to his picture of Carnap. It is clear enough, however, that Kaplan refers to the picture of Agnew in the Carnap-Agnew case, whether or not he believes that his listener is a mind reader. Contra both Reimer and Bach, unreasonable beliefs like these would appear to be referentially inert.<sup>33</sup>

Second, the success of both Reimer and Bach's proposals rely on characterizing the speaker's secondary or referential intentions in a very specific manner. In the Carnap-Agnew case, for instance, it matters that Kaplan's secondary or referential intention is thought of as an intention to demonstrate the *picture* in the direction of his pointing. Consider what would happen if instead we were to understand Kaplan as intending to refer to the *picture of a philosopher* in the direction of his pointing, or even the *picture of Carnap* in the direction of his pointing: on each of these intentions, reference fails. In order to get the case right, both Reimer and Bach assure us that the relevant intention involves the sortal 'picture', but neither offers a clear explanation for why this must be the case.<sup>34</sup> But given Kaplan's beliefs, it seems perfectly reasonable for him to intend to demonstrate not just the picture in the direction of his pointing, but also the picture of a philosopher. In fact, it seems that Kaplan might well have *all* of these intentions simultaneously; after all, he believes that all of these intentions coincide with regard to the referent that each serves to isolate. But this leaves

---

<sup>33</sup>Thanks to Jessica Pepp for helping me sharpen this objection. It is worth noting that both Reimer and Bach could try to escape this objection by claiming that, in light of her beliefs, the speaker is incompetent—and thus not really capable of saying anything at all. But this response seems off the mark: the speaker's false beliefs here don't have to do with language *per se*, but rather with human cognitive capacities. It strikes me as quite a stretch to think that people fail to count as competent speakers because they have false beliefs about humans' cognitive capacities; for one thing, many seemingly-competent speakers *do* believe in things like telepathy and clairvoyance, unfortunately. Rather, it seems far more plausible that linguistic incompetence hinges on *linguistic* misunderstanding.

<sup>34</sup>Reimer (1992) at least recognizes that some sort of explanation is required here (pp. 391–92). Unfortunately, her explanation is not actually adequate to answer this worry. What Reimer tells us is that:

[I]n cases where the demonstratum is not perceived by the speaker will the secondary (disambiguating) intention be derived from the conjunction of the primary intention and de dicto beliefs about 'the F in the range of the demonstration'. (Reimer 1992, p. 392)

The problem is that this doesn't help us choose between competing F's in context, which is the situation we are faced with. Bach (1992a,b), on the other hand, offers us no explanation at all in this regard; he merely tells in particular instances us what he takes the speaker's referential intention to be.

us with precisely the sort of intentional glut that made simple intentionalism unappealing. The first of these intentions points toward the picture of Agnew being the referent, while the latter two point towards reference failing. Without some reason for thinking that only the first of these intentions matters in context, we seem to have made remarkably little progress by adopting either Reimer or Bach's version of structured intentionalism.<sup>35</sup>

### 3.5 Smith-Jones Revisited

So far I have introduced three popular, sophisticated versions of intentionalism about demonstrative reference and argued that, even with regard to just the reference of demonstratives, we have reason to reject each of these theories. Now, I wish to briefly consider how each will apply to the Smith-Jones case. If we had any doubts about whether these theories could be salvaged in light of my arguments above, seeing how the theories apply to names should suffice to put those doubts to rest—at least on the assumption that we aspire to offer not just a semantic theory that accounts for what demonstratives mean in context, but rather a unified semantics for referential terms.

First, consider what the occurrent perception theory will have to say about the Smith-Jones case: since the speaker is occurrently perceiving only Smith, not Jones, she has a perceptually grounded intention to refer to Smith, but no such intention to refer to Jones. Thus, the speaker's use of 'Jones' is predicted to refer to Smith. But that is just plain wrong.

Next, consider King's coordination account: since the speaker intends to refer to both Smith and to Jones, reference fails. But if it just so happens that Jones, wherever he is, *is* making a hash of his raking, it looks like the speaker's utterance should be true. The coordination account looks ill-positioned to explain this fact.

Finally, consider what structured intentionalists like Reimer and Bach will have to say about the case: plausibly, the speaker has a secondary or referential intention to demonstrate

---

<sup>35</sup>Similar objections can be run *mutandis mutatis* against other versions of the neo-Gricean theory of reference—e.g. Schiffer (1981) and Neale (2004). I will not go into the details of this here.

the man in the distance, i.e. Smith. But this leads to the prediction that the speaker succeeded in referring to Smith. Once more, this looks like the wrong result.

We have thus seen that each of these three versions of intentionalism fails both with regard to demonstratives, and even worse with regard to other sorts of referential terms, like names. The question is whether we can somehow do better. I think we can.

### 3.6 The Constraint Theory of Reference

Now I turn to my positive proposal, the ‘constraint theory’. The constraint theory bears some superficial similarity to both Reimer (1991a, 1992)’s quasi-intentionalism and Bach (1992a,b)’s neo-Griceanism in that it posits that speakers *typically* go to some lengths to make clear to their listeners which object they intend to refer to. In contrast to Reimer and Bach, however, the constraint theory doesn’t *require* speakers to intend to make their referential intentions evident to their listeners in order for those intentions to fix reference in context. Rather, what the constraint theory posits is that specific referential terms, like names and demonstratives, are associated with particular *public commitments* regarding the referent. These commitments, in turn, limit the ways in which how those terms can be used to refer. In using a name or demonstrative the speaker goes on the record as being committed to the world’s being a certain way—namely, such that her intended referent bears a particular name, or is relatively distal or proximal in context. When the speaker’s intended referent fails to satisfy these properties, reference fails. It is this constraining role that explains why these terms are effective *signaling devices*: listeners understand that, so long as they are in a cooperative communicative environment, the speaker will be trying to satisfy these commitments.<sup>36</sup>

---

<sup>36</sup>The basic idea here—that different referential terms are used to signal certain things about the referent to the listener—isn’t new. Evans (1982), for instance, calls this aspect of the meaning of referential terms their ‘referential features’ (p. 311). For Evans, these referential features must be satisfied for the speaker’s reference to be ‘fully conventional’. In contrast, the present theory posits that these features must be satisfied for there to be reference at all. It is this strengthened condition that effectively allows the constraint theory to use these ‘features’ to filter multiple, conflicting intentions for just those which in fact determine reference—as

In order to get the theory fully in view, it will help to take a brief detour through Bratman (1987)'s theory of intentions. The idea will be to try to use this theory to help illuminate Grice (1989d)'s suggestion that language can productively be conceived of as *inter alia* a system for signaling our intentions to each other. That thought, in turn, underwrites most of the infrastructure of the constraint theory. What remains is, first, to specify the particular public commitments that accompany uses of names and demonstratives—and, hence, the particular sorts of signals that they can be used to generate. Then I show how, by using public intentions to *filter* confused speakers' myriad intentions, the constraint theory can account for both the Carnap-Agnew and the Smith-Jones cases.

### 3.6.1 Communicative Plans and Signals

Some intentions are relatively simple, others less so. My intention to take a sip of coffee when one is sitting in front of me is relatively straightforward. My intention to take a sip of coffee when none is in front of me is bound to be more complicated. It will consist not just of an intention to raise a cup to my mouth, but first to measure some beans, grind them, heat a certain amount of water, put the grounds in the filter, pour the water over the grounds, etc. Following Bratman (1987), I call intentions of this latter sort 'plans'. Plans are basically just intentions writ large—complex intentions that subsume other complex intentions ('sub-plans') in a particular order and with the aim of accomplishing some overall goal.<sup>37</sup>

Of special interest for our purposes are plans not just for ourselves, but those which involve other agents as well. For instance, if we are playing ultimate frisbee together and I notice that you have an angle to beat your defender to your right, then I might plan as

---

we will see shortly.

<sup>37</sup>Note that I am hardly the first philosopher of language to think that planning theory might prove significant to furthering our understanding of language use. See, for instance, Grosz & Sidner (1986), Thomason (1990), Stone (2001, 2003), Stone et al. (2003), Lewis (2012), and King (2013b).

follows: you break to your right and I will pass the frisbee to you there.<sup>38</sup> In order to carry out this plan successfully, I need to somehow convey to you that you should run quickly to your right. In other words, I need to signal my intention for you to you.<sup>39</sup> In this context, I might attempt to accomplish this by first making eye contact with you and then by gesturing in the direction I want you to run. If I'm lucky, you'll recognize what I'm trying to get you to do before your defender does. Alternatively, I could just yell "Run right!"—though this would serve to broadcast my intention for you far more widely than I would probably like.

Following Grice (1989d), I would suggest that we conceive of language as being in part a system for signaling our intentions for other agents to those agents. For instance, we can use language to signal to others our intention for them to interact with the world in certain ways ("Run right!"), or to think certain thoughts. Communication, then, is achieved when the listener recognizes the way that the speaker intends for him to act, in part on the basis of recognizing that the speaker intended for him to recognize this very intention. The listener is able to do this, the idea goes, by recognizing that the speaker's utterance was in fact a signal regarding her intentions. Thus, supposing that the speaker is indeed aiming to communicate something, she will have ample incentive to generate an appropriate signal—one that corresponds to her intentions in the right way. Call this the 'communicative signal' generated by the speaker. The plan to generate such a signal will constitute a discrete sub-part of the speaker's overall communicative plan. The question in which we are interested here is: what exactly it is that uses of names and demonstratives contribute to this communicative signal?

---

<sup>38</sup>Thanks to Hallie Liberto for suggesting this particular example to me.

<sup>39</sup>Some may be tempted to conclude from this that I cannot *really* intend for you to  $\psi$  at all, since intentions should play a causal role in generating actions and my intentions for you cannot play this role. Rather, all that I can really intend is to bring it about that you  $\psi$ . While I am unsure if this particular bit of conceptual analysis is correct, I have nothing against talking in terms of intentions to bring it about that another agent  $\psi$ 's instead of intentions that another agent  $\psi$ . What's more, nothing here hinges on whether an agent's other-directed intentions are best cashed out as straightforward intentions that another agent  $\psi$ , or whether they are better conceived of as intentions to bring it about that other agent  $\psi$ . Those who are uneasy with my terminological choice are thus invited to substitute appropriately throughout what remains.

### 3.6.2 Public Commitments and Filters

Suppose that I want to transmit a thought to you about a particular Jones, to the effect that he is  $\phi$ . To succeed in this, I need to get you to entertain a thought about that Jones, namely that he is  $\phi$ . And in order to do that, I need to make clear to you that I intend for you to entertain a thought about that Jones. Other than by making this intention evident to you, I would seem to have few other options for getting you to think this thought. I cannot simply *make you* entertain the thought that Jones is  $\phi$  via sheer force of will. So while I might try tracking down Jones and then literally pointing to his  $\phi$ -ness, the simpler route would seem to be for me to somehow signal my intention to you and to then hope that you respond in the right way.

So how might I signal to you my intention that you entertain a thought about a particular Jones? One standard way would seem to be by uttering the name ‘Jones’. More specifically, if I want to convey to you the thought that some particular Jones is  $\phi$ , I might try saying “Jones is  $\phi$ .” Of course, this utterance alone won’t serve to make evident which particular Jones I had in mind as the referent. But supposing that some particular Jones is salient in the context (perhaps he is visible, or perhaps we were recently discussing him), then this utterance should do quite nicely. That’s because, intuitively, I have signaled to you that there is *some* Jones about whom I intend to communicate, and I have relied on you to infer that—given that I have offered no indications to the contrary—this must be the Jones I intended to say something about.

This suggests that we might think about what the use of the name N contributes to the overall communicative signal as follows: a use of N indicates both that there is some object that the listener has in mind as the referent and that this object bears the name N.<sup>40</sup> Since speakers standardly generate signals of this sort in order to help communicate thoughts about objects, the listener is likely to infer, further, that the speaker intends to

---

<sup>40</sup>I set aside the hard question of what the exact conditions are for an object to bear a name in context. Below, I will stick to cases where it’s widely accepted in a community of speakers that a person  $p$  bears the name N. For discussion, see Matushansky (2006a,b), Cumming (2007), Fara (2013a), and Jeshion (2013a,b).

convey a thought about this referent—and thus that she intends for the listener to identify her intended referent in part on the basis of this signal. We needn't posit, however, that this latter aspect of the speaker's intention is explicitly signaled by her use of a name. There is simply no need for the speaker to signal her intention to help the listener identify the referent by means of her utterance; such an intention will be presupposed by the listener, so long as she takes herself to be in a cooperative communicative environment.<sup>41</sup>

Now we need to specify what these observations about signaling mean for the theory of reference. My suggestion is that, by signaling to the listener that her intended referent bears the name N, the speaker thereby publicly commits herself to her intended referent bearing that name. In other words, in using the name N, the speaker undertakes a certain normative commitment with regard to the way the world is: she goes on the record for there being something she intends to refer to, and for that that object bearing the name N. If either of these properties fail to obtain, then the speaker has *misused* the name in question. When speakers misuse names in these ways, reference fails.<sup>42</sup> It is in virtue of these public commitments that names can serve as the sorts of communicative signal they do. Listeners standardly assume that speakers are trying to live up to their public commitments—and speakers, knowing this, can use those commitments to help guide their listeners to their intended referents.<sup>43</sup>

How then to think about demonstratives? Here is a suggestion: in contrast to names, demonstratives publicly commit a speaker to her intended referent being either relatively

---

<sup>41</sup>In non-cooperative environments, on the other hand, listeners are likely to suspect that the speaker is being less than forthcoming with the information she offers about her intended referent; thus, in the very case where it might help to have a signal to indicate that one really is attempting to guide the listener to the referent, we find that the use of a referential term is ineffective as such a signal. We thus have little reason to posit that uses of referential terms play such a signaling role, as opposed to positing that this is simply a background assumption that obtains in cooperative contexts. For further discussion, see chapter 4.

<sup>42</sup>With an utterance of "N is  $\phi$ ," the speaker also presumably goes on the record for N being  $\phi$ —but this commitment is not specific to the use of the name itself. Rather, this sort of commitment attaches to uses of utterances, not to uses of names *per se*. In contrast to the sorts of commitments specifically attaching to names specifically, failure with regard to this other sort of commitment plausibly results in falsity.

<sup>43</sup>One exception might be empty names—particularly where these are commonly known not to refer. Since empty names are explicitly non-referential, however, I am not concerned with accounting for their use here.

distal or proximal, depending on whether she used the term ‘this’ or the term ‘that’.<sup>44</sup> Additionally, when uses of demonstratives are accompanied by ostensive gestures, the speaker is further committed to the referent being in the range of values indicated by her gesture. Once again, when the speaker fails to have an object in mind as the referent, or when the speaker’s intended referent fails to live up to these other public commitments, her use of ‘this’ or ‘that’ will fail to refer.

More however must be said about this notion of a *range of values* indicated by an ostensive gesture. The basic idea is that ostensive gestures are conventionalized means of directing attention to particular regions of space, and thereby to the objects and properties in those regions.<sup>45</sup> It is in this sense that I will speak of ostensive gestures as ‘indicating’ ranges of values: they ordinarily serve, in virtue of certain conventions associated with them, to make a set of objects (i.e. those located in a particular region of space) salient in context. Crucially, ostensive gestures are not, on their own at least, the right sorts of things to isolate particular objects or properties, except perhaps in some rather peculiar cases. Still, by serving as

---

<sup>44</sup>In fact, the constraints on how demonstratives can be used to refer are more complex than I will elaborate in the main text, in at least two ways. First, the proximity and distance in question here need not be physical proximity and distance. Rather, these scales seem equally applicable to time or place in story or argumentative structure. Surprisingly, the question of how exactly to characterize the bounds of these notions of *grammatical* distance or proximity (I use the term ‘grammatical’ here to mirror the notion of *grammatical gender*—a notion that extends well beyond objects literal sex) has received scant attention in the empirical literature. Grammatical distance and proximity would seem to be similarly bound to specific languages, with Japanese and Spanish, for instance, admitting of a three-way distinction (‘sore’/‘kore’/‘kare’ and ‘ese’/‘este’/‘aque’, respectively), in contrast to the relatively impoverished two-way distinction found in English. The only work of which I am aware that even comes close to addressing this issue focuses primarily on when bare demonstratives can appropriately be used in context, as opposed to definite descriptions or indexicals like ‘it’ (cf. Gundel et al. 1993, 2003, Wolter 2006). Note also that there are plausibly languages with demonstratives that fail to encode anything about relative distance. Lao, for instance, seems to be one such language (cf. Enfield 2003). German is plausibly another. Second, bare demonstratives (in English, at least) come with a complex animacy constraint: they can be used to felicitously refer to animates only when flanked by predicate nominals that specifically entail animacy. For instance, uttering “That’s funny” while pointing at and intending to refer to David Kaplan is highly marked, whereas uttering “That’s my advisor” under similar circumstances is not. See Wolter (2006) for a discussion of ‘introductory uses’ of demonstratives that touches on this issue.

<sup>45</sup>While the conventionality of ostensive gestures might not be immediately apparent, variance in how such gestures function, and what sorts of gestures are used in ostentation is well-attested. For instance, as was noted in chapter 2, some socio-linguistic groups use lip pursing in much the same way that most readers are probably accustomed to gesturing with an extended finger (cf. Key 1962, Sherzer 1972, 1983, 1993, Poyates 1983, Feldman 1986, Enfield 2001, Wilkins 2003). And, indeed, there is both far more complexity and far more conventionality to gesturing with one’s hands than one might initially imagine (cf. Haviland 2003).

conventionalized cues that one should direct one's attention to a particular region of space, ostensive gestures are precisely the right sorts of things to indicate ranges of values.<sup>46</sup>

In summary then, the constraint theory proposes to treat the reference of names and demonstratives, in context, along the following lines:

NAMES        The use of a name  $N$  refers to  $o_1$  when (i) the speaker intends to refer to  $o_1$ , (ii)  $o_1$  bears the name  $N$  in context, and (iii) there is no object  $o_2$  such that  $o_1 \neq o_2$  and both (i) and (ii) also hold for  $o_2$ . Otherwise, it fails to refer.

DEMONSTRATIVES        The use of a demonstrative refers to  $o_1$  when (i) the speaker intends to refer to  $o_1$ , (ii)  $o_1$  is relatively distal ('that') or proximal ('this') in context, (iii)  $o_1$  falls within the range of values indicated by the speaker's ostensive gesture, supposing there is one, and (iv) there is no object  $o_2$  such that  $o_1 \neq o_2$  and both (i)–(iii) also hold for  $o_2$ . Otherwise, it fails to refer.

One aspect of the constraint theory worth emphasizing is that, in contrast to most previous versions of intentionalism, the theory is designed to allow for reference to succeed even in cases where speakers have multiple, non-equivalent intentions to refer. In effect, when speakers have multiple, non-equivalent intentions to refer—that is, when they intend to refer both to  $o_1$  and to  $o_2$ , where  $o_1 \neq o_2$ —the public commitments of the particular referential term they use serve effectively to *filter* these intentions for just those intentions which satisfy these commitments. So long as the speaker has at least one referential intention that satisfies these commitments, and so long as she has no other, non-equivalent referential intentions that also satisfy these commitments, then her use of the relevant term will succeed in referring according to the constraint theory. We will return to the significance of this filtering effect shortly, when we turn back to the cases of referential confusion with which we began.

Before that, however, it will help to illustrate the constraint theory by means of a few simpler examples. Consider a case in which I have just one referential intention—to refer to Smith—but for some reason I utter the name 'Jones'. Perhaps I mistakenly believe that

---

<sup>46</sup>Some have been tempted to think of gestures, in virtue of their conventionality, as having *meanings* in context—and thus to speak of them as 'denoting' ranges of objects (cf. Lascarides & Stone 2009). Since I don't need anything so strong here, I will eschew this possibly confusing terminology. Nonetheless, I am sympathetic to the idea of associating ranges of values with gestures as their denotations.

Smith is named ‘Jones’, or perhaps I have just misspoken. According to the constraint theory, reference fails here because I have no Jones in mind to whom I intend to refer.<sup>47</sup> Likewise, consider a case in which I intend to refer to my nose with a use of ‘this’, and I further intend to point to my nose in the process of making my utterance, but for whatever reason I end up pointing directly at my left foot. Here again, the constraint theory predicts that reference fails—as once more seems appropriate.<sup>48</sup>

This constraint theory thus offers a unified account of linguistic reference in the sense that it posits that reference, in context, depends on speakers’ intentions as constrained by, and filtered through, their public commitments. The theory is less unified than simple intentionalism in that it posits that different sorts of referential terms are associated with different sorts of public commitments. This disunity is hardly a bad thing, however. After all, while names and demonstratives are both referential terms, they cannot easily be substituted for each other in context while retaining both reference and felicity. Other versions of intentionalism offer no clear way of accounting for these differences in meaning; in contrast, such an account is exactly what the constraint theory provides.<sup>49</sup>

---

<sup>47</sup>Some may find this counterintuitive in the limit case, where there really is just one individual named ‘Jones’. Since such cases are true outliers, we could append the theory to handle them separately if necessary. However, I am skeptical that we should. A speaker who really has *no one* in mind named ‘Jones’ to whom she intends to refer with her use of the name ‘Jones’—not even the minimal ‘the fellow I’ve heard of named ‘Jones’—strikes me as plausibly failing to refer. The problem is that she isn’t using the name in a way that is properly synced with the actual Jones; she hardly seems to be using the name as a name for *him*.

<sup>48</sup>We must be careful, however, to rule out unintentional ticks as counting as gestures. Gestures, I take it, are intentional acts. Even if one’s tick happens to *look* precisely like a pointing gesture, it should not be understood as such. Hence, it will not play a constraining role on reference—though listeners who aren’t in the know regarding a speaker’s ticks will hardly be culpable for thinking that she has gestured in such circumstances. What frustrates reference, I take it, are intentional gestures that somehow go wrong.

<sup>49</sup>Crucially, it will not do (at least on standard theories) to simply associate the use of a name or demonstrative with a particular *presupposition*. This is because presuppositions are typically thought to attach to antecedently determined referents, causing infelicity (or even reference failure) when they are violated; they have not standardly been understood to be the sorts of things that can help to actually *determine* what the reference is in context. The notion of a public commitment invoked here, on the other hand, is introduced in order to play precisely this positive role in *determining* reference—not just determining reference failure or infelicity. See Cumming (2013) for a related use of the notion of a public commitment. Note as well that these observations complicate the move that many more standard intentionalists are likely to reach for when it is pointed out that, *prima facie* at least, their theories offer no way of explaining the obvious fact that ‘this’ and ‘that’ actually *mean different things*. Supplementing standard versions of intentionalism with a presuppositional component will indeed serve to answer this latter worry, but it is no panacea. In particular,

### 3.6.3 A Guide to Handling Confusion and Absence

Now it is time to deploy the constraint theory in order to account for reference in the cases of confusion with which we began. I begin with the Carnap-Agnew case. In that case, Kaplan intends, in virtue of his confusion, to refer both to the picture of Carnap and to the picture of Agnew. However, since he also points behind himself, the constraint theory predicts that his use of ‘that’ publicly commits him to the referent being behind him in the context. Thus, only one of his intentions thus passes through the filter associated with his use of ‘that’ in conjunction with his pointing: his intention to refer to the picture of Agnew. That’s because only the picture of Agnew is in the set of things indicated by Kaplan’s pointing; the picture of Carnap, having been removed, is not. So, according to the constraint theory, only Kaplan’s intention to refer to the picture behind him is relevant, and thus Kaplan refers to the picture of Agnew.

In the Smith-Jones case, the speaker intends both to refer to the man in the distance raking leaves (i.e. Smith), and to a particular Jones whom she has mistaken for Smith. Since only Jones bears the name ‘Jones’, however, only the latter intention passes through the filter associated with the speaker’s use of the name ‘Jones’. Thus, according to the constraint theory, it is only the speaker’s intention to refer to Jones that matters for fixing reference here.<sup>50</sup>

For good measure, let us also consider how the constraint theory might address King (2013a)’s absent skier case, even though this is a case of deferred reference rather than confusion. Recall that, in that case, the speaker pointed to an empty chair and used the gendered pronoun ‘he’ with the intention of referring to Glenn. The idea was that the speaker could

---

it will not serve to explain the range of linguistic behavior we observe in confusion cases.

<sup>50</sup>There is, in fact, an even harder case in the vicinity, one that has somehow gone unnoticed to this point. This is the ‘Jones-Jones’ case. In this variant of Kripke’s case, the man raking leaves is, by sheer coincidence, also named ‘Jones’. On the constraint theory, the speaker’s use of ‘Jones’ fails to refer in this version of the case—since both of the speaker’s intentions pass through the filter associated with her use of the name ‘Jones’, and since these two intentions are non-equivalent. This strikes me as a plausible thing to say about the case. I grant, however, that judgments on this case are likely to be murky.

succeed in referring to Glenn in such a scenario in spite of his absence. Assuming that gendered pronouns refer in a similar way to how demonstratives do—that is, in accordance with the gestures that often accompany them—this case represents a problem for the constraint theory as introduced above. Specifically, since Glenn is absent from the scene, it seems that he cannot be in the range of values indicated by the speaker’s gesture. But in that case the constraint theory should predict that reference fails.

As the theory stands, this case does indeed constitute a counterexample to it. However, the constraint theory can be easily modified so as to accommodate not just this case, but other cases of deferred reference as well.<sup>51</sup> The crucial thing to notice is that such cases—or at least those involving demonstratives—all involve reference to an object that is somehow represented by an object in the range of values that the speaker indicates by means of her gesture. The simplest cases will involve objects like pictures and photographs, which, by their very nature, are well-suited to represent other objects. What King’s case suggests, I take it, is that there are more ways for objects to represent other objects, in context at least, than just by depicting them. Rather, the fact that an object is habitually associated with a particular person would seem to suffice for the former to represent the latter, at least in certain circumstances. Deferred reference is thus not wholly unconstrained—it requires, among other things, that there be *some* object standing in *some sort of* representational relationship to the referent. Still, deferred reference is clearly significantly less constrained than more standard, non-deferred reference.<sup>52</sup>

Having noted this relationship between deferred reference and objects indicated, we are now in a position to modify the reference clause for demonstratives. This becomes:

DEMONSTRATIVES\*      The use of a demonstrative refers to  $o_1$  when (i) the speaker intends to refer to  $o_1$ , (ii)  $o_1$  is relatively distal (‘that’) or proximal

---

<sup>51</sup>The modifications suggested here will in fact largely parallel those considered earlier, in subsection 3.4.3, for Reimer (1991a, 1992)’s quasi-intentionalism.

<sup>52</sup>Just how unconstrained deferred reference turns out to be will depend on just what sorts of properties suffice to allow for one object to represent another in context. For the beginnings of an investigation into the bounds of this and other forms of metonymy, see Hobbs (1990) and Nunberg (1995).

(‘this’) in context, (iii) either  $o_1$  falls within the range of values indicated by the speaker’s ostensive gesture, supposing there is one, or else  $o_1$  is represented by some object or property  $o_2$  and  $o_2$  falls within the range of values indicated by the speaker’s gesture, and (iv) there is no object  $o_3$  such that  $o_1 \neq o_3$  and both (i)–(iv) also hold for  $o_3$ . Otherwise, it fails to refer.

This modified understanding of the public commitments accompanying uses of demonstratives allows the constraint theory to account for King’s empty chair case: since Glenn is represented by the empty chair, and since the empty chair falls within the range of values indicated by King’s gesture, King’s intention to refer to Glenn satisfies the public commitments associated with his use of ‘that’. Deferred reference thus poses no deep threat to the constraint theory. What cases of deferred reference like this one do serve to illustrate is that—even when uses of demonstratives are accompanied by ostensive gestures—the constraints on successful reference aren’t as strict as we might initially have thought.

The constraint theory thus has a good claim to be simpler, more accurate, and significantly wider in scope than any of its predecessors. It adequately accounts for referential confusion involving not only the use of demonstratives, but also names. And, in fact, it promises to extend further still. What’s more, it is motivated by a straightforward and plausible claim: that in order to successfully transmit thoughts about particular objects, speakers need some way of indicating to their listeners which objects they are trying to talk about. The constraint theory represents one way of translating this basic thought into a theory of reference. In particular, the constraint theory proposes that, while speakers’ mental lives are, often at least, deeply messy, this messiness needn’t undermine the referential efficacy of all of those speakers’ referential intentions.

### 3.7 Conclusion

The primary goal of this chapter was to make the case for the constraint theory *vis-a-vis* other forms of intentionalism about reference. By viewing uses of names and demonstratives as publicly committing the speaker to her intended referent bearing certain properties, I argued,

the constraint theory can account for the linguistic reference of names and demonstratives in a parallel manner—even in cases of referential confusion. What’s more, that theory looks well-positioned to account for a variety of other referential terms as well: gendered pronouns, complex demonstratives, and even referential uses of definite descriptions.

That said, the constraint theory does bring with it one potentially serious drawback: according to that theory, speakers can successfully use names and demonstratives to refer even in circumstances where they take it to be very unlikely, or even impossible, for their listeners to figure out what they are talking about. Such cases will be communicatively sub-optimal, but that, according to the theory, is neither here nor there from the point of view of reference. I suspect that many—motivated by something like the thought that semantic theory earns its keep by helping to explain successful communication—will find this result unpalatable. In the next chapter, I argue at length that this reaction is unwarranted. Contrary to this suggestion, there is good reason to think that gaps can arise between what speakers intend to refer to and what they intend for their listeners to recover in context as the referent. This is a phenomenon I call ‘sneaky reference’.

By way of conclusion, I want to briefly reflect on the phenomenon of referential confusion and its relation to semantic content more generally. Both the Carnap-Agnew and Smith-Jones cases exhibited an interesting property with regard to confusion: although the speaker was confused, that confusion didn’t manifest in the truth-conditions of her utterance. Those, in contrast, were clear enough. This observation might seem to speak against intentionalist theories of reference, or even of meaning, more generally. The problem is that intentionalism can seem to be motivated by the following idea: standardly, we use language to *express* our thoughts. Intentionalism, then, might be characterized as the view that this expression manifests directly in fixing the truth-conditions of our utterances. With respect to referential terms, what our thoughts are about should thus be reflected in what those terms refer to in context. In cases of confusion, however, we saw how linguistic reference can pull apart from what the speaker is thinking about; the speaker’s thought wasn’t about any one object, she

managed to talk about a particular object nonetheless.

What should we make of this? What this suggests, I take it, is that this way of motivating intentionalism is getting something subtly wrong. While our utterances may help to transmit our thoughts, this doesn't require that our utterances directly express those thoughts. One way of probing this suggestion is by considering whether we can transmit confused thoughts by means of utterances with determinate truth-conditions. Quite possibly, we can. Consider a variant of the Carnap-Agnew case in which the listener knows that there is ordinarily a picture of Carnap hanging behind Kaplan's desk, and where he, like Kaplan, has failed to notice that this has been replaced by a picture of Agnew. Or consider a version of the Smith-Jones case where the listener mistakes Smith, who is raking leaves in the distance, for the same Jones that the speaker mistakes him for. In both these cases, it looks like the speaker's utterance should result in the listener's entering a state of confusion paralleling the speaker's own confused state. This, I would suggest, is precisely what it would mean to transmit a confused thought.<sup>53</sup> If that's right, then it looks like we can indeed transmit thoughts that are truth-conditionally distinct from the contents expressed by the utterances used to transmit them—even in cases of perfectly literal speech and communication.

But if the truth-conditional content associated with an utterance in context doesn't necessarily reflect the speaker's underlying thought, what sort of role *does* such content play in an overall theory of language use? According to one popular theory, truth-conditional content helps to explain how successful communication is achieved: speakers first translate their thoughts into truth-conditionally equivalent sentences. Then, they verbalize those sentences. The listener's job is to decode these utterances, which they do precisely when they manage to recover the truth-conditions of those utterances. Such recovery, in turn, constitutes the transmission of a thought—and, thus, communicative success.<sup>54</sup> In cases of

---

<sup>53</sup>For discussion of how one might try to assign determinate contents to confused thoughts, see Lawlor (2005, 2007). If successful, this project would thus offer another, more formal, way of characterizing what it means to transmit a confused thought: prompting the listener to think a thought with either the same content as the speaker's own confused thought, or else with some analogous content.

<sup>54</sup>See Dummett (1976) for a classic defense of this sort of approach to meaning. For a more recent

referential confusion, we see how truth-conditional content can fail to reflect the content of the speaker's thought. This serves to undermine this neat story of uptake-as-understanding—or at least as the sort of understanding that suffices for the transmission of a thought. Thus, we cannot simultaneously endorse both the constraint theory and this picture of the role of truth-conditional content in a theory of language use.

Note, however, that in introducing the constraint theory, we already implicitly sketched an alternative account of communicative success: communication succeeds when (i) the speaker aims to transmit a particular thought by means of her utterance, (ii) the listener recovers that very thought, and (iii) he does so on the basis of interpreting the speaker's utterance.<sup>55</sup> Nowhere here is it necessary to appeal to the truth-conditions of the utterance itself. The question, then, is what truth-conditions might be good for if they are in fact unnecessary for explaining communicative success. Here is one suggestion: truth-conditional content tracks *one core element* of the various sorts of responsibilities that speakers bear for their utterances. Plausibly, truth-conditions track a privileged sort of responsibility—responsibility for the content strictly put forward by an utterance.

If this is right, then we are licensed to reason as follows. Assuming that our judgments on truth and falsity are reasonably accurate, and assuming further that this core type of responsibility obtains in virtue of a practice we have developed for holding each other accountable for our utterances, then this practice has developed, to some extent at least, so as to resist speaker confusion. This is an interesting conclusion, but perhaps not an altogether surprising one. After all, it would seem to be quite useful to have some way of classifying the things we say as accurate or inaccurate, independent of the various epistemic imperfections

---

discussion, see King & Stanley (2005). Even if, per classical Gricean theory, we complicate this picture by adding that sometimes communicative success relies on recovering more than just the truth-conditions of the utterance, the bulk of this picture will remain unchanged. That's because, according to Gricean theories of pragmatic inference, the truth-conditional contents of utterances serve, at least in principle, as a key input for subsequent pragmatic reasoning (Grice 1989a, pp. 30–31). Thus, they must still be recovered, or at least be recoverable, as part of the overall process of successful communication.

<sup>55</sup>Note that, due to shifts in perspective, this will not always be the thought that the speaker herself began with—or even one that she is herself capable of thinking. For discussion, see Heck (2002).

that we commonly suffer from, and independent of the various effects that those imperfections can have on what we might attempt to communicate with a particular utterance. Such judgments might, for instance, help to inform our assessments of particular speakers' reliability—or perhaps reliability in some specific domain.

To accept something along these lines, however, is to effectively assign truth-conditions to a very different theoretical role than the one with which we began: what truth-conditions help to explain, on this sort of picture, is not successful communication in context, but rather our practice of assessing each other as reliable or unreliable speakers, independent of any particular communicative aims we might have in context. In non-confused cases, communication may well take place via the listener's recovering the truth-conditional content expressed by the speaker's utterance. But this is just one sort of case; the fact that content and communication co-vary in such cases is not enough to ground that content's being assigned an explanatory role in the theory of communication *simpliciter*. Rather, given that it arises even in cases of confusion—where its recovery would not result in the successful transmission of the speaker's thought, and where successful transmission of a thought does not depend on the recovery of truth-conditional content—it seems that truth-conditional content may be best-suited to play a different explanatory role than this one.

Thus, while the thought that truth-conditional content earns its keep by helping to explain successful communication is certainly an appealing one, it looks like we have reason to reject it. On that picture of the relationship between semantics and communication, the robustness of truth-conditional content in the face of speaker confusion would be highly unexpected. If, on the other hand, what truth-conditional content reflects is our practice of holding each other accountable for the things we say, then this robustness is exactly the sort of quality we might expect for truth-conditional content to exhibit. All else equal, and given our epistemic limitations as human beings, a set of linguistic conventions that generates such robustness will be far more useful towards this end than a similar set of conventions, but one that cannot tolerate cases of referential confusion, would be.

# Chapter 4

## Sneaky Reference

*Await not in quiet the coming of the horses, the marching feet, the armed host upon the land. Slip away. Turn your back. You will meet in battle anyway. O holy Salamis, you will be the death of many a woman's son between the seedtime and the harvest of the grain.*  
*The Delphic Oracle—circa 480 BCE*

### 4.1 Introduction

In the last chapter, we explored several versions of sophisticated intentionalism, including both the neo-Gricean and the constraint theories of reference. While the arguments developed in that chapter should be sufficient to justify a preference for former, the latter theory merits further consideration nevertheless. The reasons for this are two: first, the neo-Gricean theory of reference is essentially a specific application of the broader, and still widely-accepted, Gricean theory of meaning. Thus, we might reasonably hope to learn more about the virtues and drawbacks of this broader theory by investigating its implications for the theory of reference. Second, the primary objection to the neo-Gricean theory offered in chapter 3 might seem to be a largely technical objection—one that can perhaps be overcome with sufficient ingenuity. After all, that objection hinged on speakers with aberrant beliefs about the world. Perhaps there is some way to rule out the effects of these beliefs. Or

perhaps the neo-Gricean will simply prove willing to accept that odd things happen when strange enough speakers are involved.

This chapter will argue that a deeper objection lurks in the background here—one that also serves to undermine not just Grice’s own theory of meaning, but any such theory that depends on the basic claim we find at the heart of the Gricean theory. That claim runs roughly as follows: the specific intentions that serve to fix meaning in context are intentions for the listener to react in a particular way, based on his recognition that this is what the speaker intends for him to do. The present chapter will argue that, in endorsing a claim along these lines, what both Grice and, subsequently, a whole host of neo-Griceans have done is to mistake what is plausibly a *good making* feature of certain sorts of speech acts—i.e. cooperativeness—for a constitutive condition on meaningfulness.

The problem is that we are capable of doing plenty of noncooperative things with our words and utterances while still meaning things by them. In particular, we are plausibly capable of *deceiving* each other not just about how the world is, but about what our utterances themselves mean. If that’s right, then we must be capable of using an utterance to mean *p* while simultaneously intending for our audience to think that we mean *q*. That, in turn, means that intentions for the listener to react in a certain way, on the basis of the listener’s recognitions of that very intention, cannot be what determines the meaning of utterances in context. In other words, Grice and others have assumed that meaning-fixing intentions must either be, or at least aim to be, *transparent* to the listener. In this chapter, I will make the case for the potential for such intentions to be knowingly and intentionally *opaque*.

To make that case, I will focus once more on a set of examples drawn from the theory of reference. These examples are particularly useful for our purposes because, as we saw in the last chapter, the question of how to apply the Gricean theory of meaning to the phenomenon of reference has already been worked out in significant detail. What’s more—and again as we saw in the last chapter—this neo-Gricean theory of reference was superior in certain respects to many of its competitors: it is broad in its scope, capable of handling at least the

classic case of speaker confusion (i.e. the Carnap-Agnew case), and offers a detailed picture of the exact sorts of intentions that fix meaning on particular occasions. I argued above that my own constraint theory was able to match the neo-Gricean with regard to each of these virtues. In this chapter, I will buttress the case in favor of the constraint theory by showing that it has an additional virtue as compared to the neo-Gricean theory: it can account for what I will call ‘sneaky uses’ of referential terms as well.

Here is the plan for what is to come: in section 4.2, I introduce the Gricean theory of meaning in more detail. Then, in section 4.3, I briefly re-introduce how this theory has been applied to referential terms. In section 4.4, I pause to explain why either referential intentions or something very similar will be required to explain what names refer to in context. In section 4.5, I demonstrate how the reference of both names and demonstratives can violate Gricean reflexivity, specifically in cases of what I call ‘sneaky reference’. In section 4.6, I consider and ultimately reject several defensive maneuvers the Gricean might try to avail herself of. In section 4.7, I explain in more detail the ways in which the constraint theory parallels the neo-Gricean account, and also how it departs from it. Then I show how it is these points of departure that allow the constraint theory to account for cases of sneaky reference. Finally, in section 4.8, I conclude, first, by showing that the phenomenon of ‘sneaky meaning’ extends beyond just referential terms, and then by considering what this tells us about the relationship between meaning and responsibility.

## 4.2 The (neo-)Gricean Theory of Meaning

Grice (1989d) proposed that, in order for an utterance to count as meaningful, it must be accompanied by a particular sort of intention to use that utterance cooperatively. More specifically, for Grice, a meaningful utterance must be accompanied by what he calls a ‘reflexive intention’: that is, an intention to produce a particular effect in one’s listener—paradigmatically, prompting him to believe  $p$ —partly in virtue of that listener recognizing

that this is what the speaker intends for him to do (Grice 1989d, p. 220).

Here is how Grice puts the idea:

“ $x$  meant something” is (roughly) equivalent to “Somebody meant<sub>NN</sub> something by  $x$ .” (Grice 1989d, p. 220)<sup>1</sup>

This raises the question of what is required for someone to mean something by  $x$ . Here again is Grice, first from “Meaning” and then from his later “Utterer’s Meaning and Intentions”:

“ $A$  meant<sub>NN</sub> something by  $x$ ” is (roughly) equivalent to “ $A$  intended the utterance of  $x$  to produce some effect in an audience by means of the recognition of this intention.” (Grice 1989d, p. 220)

“ $U$  meant something by uttering  $x$ ” is true iff, for some audience  $A$ ,  $U$  uttered  $x$  intending:

- (1)  $A$  to produce a particular response  $r$
- (2)  $A$  to think (recognize) that  $U$  intends (1)
- (3)  $A$  to fulfill (1) on the basis of his fulfillment of (2). (Grice 1989b, p. 92)

So, according to Grice, for an utterance or expression to be meaningful on a given occasion, it must be used by a speaker with a certain sort of intention, a reflexive intention. That is, it must be used by a speaker with the intention that it produce a particular response in the listener—in part by means of his recognizing this very intention.

For Grice, paradigm reflexive intentions are aimed at producing beliefs in the listener (Grice 1989d, p. 219). Subsequent neo-Griceans, however, have often taken issue with this aspect of Grice’s account. For instance, Searle (1969), Armstrong (1971), Schiffer (1972), Bennett (1976), Vlach (1981), and Neale (1992) have all argued that it is not beliefs that speakers aim to elicit in uttering sentences. Rather, what speakers aim to achieve is to prompt the listener: to entertain  $p$  (Neale 1992), to take themselves to have evidence that the speaker believes  $p$  (Armstrong 1971), to take themselves to have some evidence in favor

---

<sup>1</sup>In Grice’s notation, ‘means<sub>NN</sub>’ stands for the sort of meaning that properly attaches to bits of language, via ‘non-natural’ conventions. Non-natural meaning thus stands in contrast to the sort of ‘natural meaning’ that can arise via correlations of objects or properties with other objects or properties, e.g. smoke ‘meaning’ fire (Grice 1989d, p. 214). Non-natural meaning will serve as our sole concern in this inquiry.

of  $p$  (Bennett 1976), to interpret the speaker's utterance as meaning  $p$  (Searle 1969), or to take the speaker to have transparently committed herself to  $p$  (Vlach 1981). What these views share in common is the claim that the speaker aims to get the listener to  $\phi$  in part by means of recognizing that this is what the speaker was trying to get him to do, via her utterance. Call this the 'kernel' of the Gricean theory of meaning. It is this kernel that will serve as my target below, rather than Grice's particular commitment to reflexive intentions being aimed at eliciting beliefs. That said, for ease of exposition, I will focus primarily on Grice's own version of the theory; it should be clear enough how to extend the arguments so as to address these various modifications to the original Gricean view.

Note that, so far, Grice (and, likewise, these various neo-Griceans) has really only provided a theory of *meaningfulness*, not a theory of meaning. To provide an adequate theory of what particular expressions mean on particular occasions, we are going to have to start filling in the sorts of responses  $r$  that those expressions can be used to elicit, and then clarifying how these responses relate to specific contents or other sorts of meaning. In the case of indicatives, Grice suggests that the desired effect is to prompt the listener to believe some content  $p$  (Grice 1989d, p. 219).<sup>2</sup> This, in turn, offers a basis for associating particular uses of indicative sentences with particular contents  $p$ . With regard to sub-sentential expressions, however, things are trickier; it is often highly unclear just what reaction the speaker is trying to elicit on the basis of her use of a single word, as opposed to her use of the whole of the sentence within which that term is embedded.

A striking exception is the class of referential terms. With respect to these, there seems to be a relatively straightforward answer to the question of what the relevant  $r$  should be: referential terms are standardly used to prompt the listener to identify a particular object as the referent. This thought has been taken up by a number of neo-Griceans in order to demonstrate the usefulness of the Gricean theory of meaning to the project of lexical

---

<sup>2</sup>The various neo-Griceans canvassed would thus have to modify this to: entertaining the belief that  $p$ , taking oneself to have evidence in favor of  $p$ , etc.

semantics—and, in particular, to the theory of reference (cf. Schiffer 1981, Bach 1992a,b, Neale 2004). As we saw in the last chapter, the resultant theory is in fact reasonably successful. In the next section, I will briefly review both that theory and its virtues.

### 4.3 The neo-Gricean Theory of Reference

The neo-Gricean theory is characterized by a pair of related claims: (i) terms like demonstratives and pronouns refer to the object isolated by the speaker's *referential intention*. And (ii) referential intentions must exhibit the Gricean reflexive form. That is, properly speaking, these are intentions for the listener to identify some object as the referent—in a particular manner, and on the basis of recognizing this very intention. Here again is Bach's summary of the position:

The relevant intention here, the specifically referential one, is part of a communicative intention, a reflexive intention whose distinctive feature is that 'its fulfillment consists in its recognition' by one's audience, partly by supposing (in Gricean fashion) that the speaker intends his intention to be recognized...[A] referential intention isn't just any intention to refer to something one has in mind but is the intention that one's audience identify, and take themselves to be intended to identify, a certain item as the referent by means of thinking of it a certain identifiable way. (Bach 1992a, p. 144)<sup>3</sup>

Crucially for our purposes, and as should be apparent from this passage, neo-Griceans like Bach require that referential intentions aim at somehow making the referent evident to the listener. In other words, according to Bach, referential intentions are required to aim at achieving a particular sort of transparency.

---

<sup>3</sup>An earlier, and somewhat more technical, version of this same idea can be found in Schiffer (1981):

*S* refers to *x* in uttering *u* if  $(\exists A)$  in uttering *u* *S* means that *x* is such and such, or that *A* is to make it the case that *x* is such and such. (Schiffer, 1981, p. 69)

Here, '*S*' is the speaker, '*x*' is an object (specified, for Schiffer, *de re*), '*A*' is an audience, and 'means' is being used in a technical sense that requires a reflexive intention on the part of the speaker. Another variant of this theory can be found in Neale (2004) and is discussed in greater detail in section 4.6 below.

One advantage to this sort of theory is the amount of detail it offers with regard to the nature of referential intentions—particularly in contrast to simpler forms of intentionalism. A second, and perhaps more significant, advantage is that the neo-Gricean theory of reference makes accurate predictions in a range of hard cases. For example, as we saw in the chapter 3, the neo-Gricean theory is capable of explaining Kaplan (1978a)’s classic ‘Carnap-Agnew’ case. In that chapter I largely glossed over the details of how the neo-Gricean can explain this case. Now it is time to explain this more precisely.

Recall how Kaplan (1978a) originally introduced this case:

Suppose that without turning and looking I point to the place on my wall which has long been occupied by a picture of Rudolf Carnap and I say:

- (4) That is a picture of one of the greatest philosophers of the twentieth century.

But unbeknownst to me, someone has replaced my picture of Carnap with one of Spiro Agnew...I have said of a picture of Spiro Agnew that it pictures one of the greatest philosophers of the twentieth century. (Kaplan 1978a, p. 239)

The goal then is to explain how an intentionalist theory of demonstrative reference can explain the falsity of (4).

In order to handle the case, the neo-Gricean will claim that Kaplan intends for his use of ‘that’ to refer to the picture of Agnew in virtue of his intention that the listener identify the referent in something like way *w*: look in the direction of my (Kaplan’s) gesture and identify the only picture in the vicinity of that gesture as the referent. Of course, Kaplan also believes that this *w* will serve to identify his picture of Carnap as the referent. As it happens, however, *w* only serves as a way of identifying the picture of Agnew as the referent in this context. Thus, according to the neo-Gricean, it is only Kaplan’s intention to refer to the picture behind him, the picture of Agnew, that fixes reference here—since only Kaplan’s intention to refer to that picture exhibits the sort of reflexive form required to count as a referential intention. On the neo-Gricean theory, it is just this part of Kaplan’s overall mental state that constitutes a genuine referential intention, as opposed to something that

Kaplan might confusedly believe to be a referential intention (Bach 1992b, pp. 296–97). And, according to neo-Griceans like Bach, it is only genuine referential intentions that matter for fixing reference in context (Bach 1992a, p. 143).<sup>4</sup>

More accurately, the neo-Gricean actually makes two separate, although related, predictions here: first, what Kaplan *speaker refers* to in this context is the picture of Agnew. In other words, what *Kaplan* refers to in this context is that picture. Second, what Kaplan’s use of ‘that’ refers to in this context is also the picture of Agnew.<sup>5</sup> According to the neo-Gricean, both of these follow from the fact that what Kaplan intended for his listener to identify as the referent was the picture in a certain demonstrated location—something which turned out to be a picture of Agnew.<sup>6</sup> In fact, both of these predictions seem quite reasonable: Kaplan plausibly did intend to refer to the picture behind him, the picture of Agnew, and his use of ‘that’ plausibly referred to that picture as well.

Unfortunately, while the neo-Gricean theory generates what is plausibly the right prediction with regard to the Carnap-Agnew case, it stumbles on a different set of cases—cases involving a particular sort of deception. Since some of the clearest instances of this phenomenon involve names, I must first explain why there is reference-fixing problem for names

---

<sup>4</sup>One might wonder whether Kaplan could also be open to his audience identifying his intended referent in way  $w^*$ : identify my picture of Carnap as the referent, on the basis of recognizing that this is the picture that ordinarily hangs behind me (and which is presumably still there). This possibility would be particularly salient in the case where the listener is another philosopher, one who is both well-acquainted with what Carnap looked like and with the standard contents of Kaplan’s office. The problem is that, were Kaplan indeed open to his listener identifying the referent by means of  $w^*$ , then both his intention to refer to the picture of Carnap and his intention to refer to the picture of Agnew would exhibit the required Gricean reflexive form. That, in turn, would serve to undercut the neo-Gricean’s explanation of the case. Unfortunately, even where Bach addresses this case most explicitly (i.e. Bach 1987, pp. 182–86; Bach 1992b, pp. 296–97), he fails to consider this possibility. Nor am I aware of any other neo-Griceans who address this worry. While this strikes me as a fairly serious problem for the neo-Gricean, I set it aside in what remains.

<sup>5</sup>To be clear, for neo-Griceans these two types of reference will never diverge. There is thus a sense in which neo-Griceans might be tempted to resist the thought that these constitute separate predictions. However, it will help to treat them as such here, since it is standardly assumed that truth-conditions are determined by what expressions refer to in context, rather than what speakers refer to (cf. Kripke 1977).

<sup>6</sup>This latter sort of reference, pertinent to the determination of the truth-conditions of utterances in context, often goes by the name ‘semantic reference’ (cf. Kripke 1977). I will tend to shy away from that term in the main text for the following reason: Bach, for one, denies that reference fixed by anything other than purely linguistic rules is, properly speaking, semantic (cf. Bach 1992a, 2005, 2006). Still, according to Bach, non-semantic reference can help to determine the truth-conditions of utterances in context.

paralleling the one for demonstratives. Due in large part to the ways that philosophers have typically approached the semantics of names, this problem may seem far from obvious.<sup>7</sup>

## 4.4 Names and Referential Intentions

Philosophers tend to think about names in the following way: ‘Jones’ refers to Jones.<sup>8</sup> The interesting task is to try to get a grip on how this reference relation gets established between the name and the object named, and how it can be passed along from one speaker to another.<sup>9</sup> Our starting point here will be rather different. In particular, we are going to give up the convenient fiction that, in any given speech community, there is only one individual who bears a given name. Thus, we will no longer be able to say, simply, that ‘Jones’ refers to Jones; the problem is that there is more than one Jones to whom a use of ‘Jones’ might refer. This leaves us with the following question: in virtue of what does a *particular use* of the name ‘Jones’ refer to some *particular* Jones? Like the demonstratives ‘this’ and ‘that’, it seems that the name ‘Jones’ is somehow *used by speakers* to refer to different objects on different occasions—specifically, to objects that bear the name ‘Jones’.<sup>10</sup>

The neo-Gricean theory of reference offers a ready-made answer to this question: what makes a particular use of a name refer to what it does is that the speaker intends for the listener to identify a particular object as the referent, and that she has in mind some particular way for the listener to identify that object as the referent (in part, presumably,

---

<sup>7</sup>Note that, in chapter 3, I largely took it for granted that there is a reference fixing problem with respect to names. Now it is time to introduce the problem more explicitly.

<sup>8</sup>See, for instance, Kripke (1977, p. 263). Ryle (1949) earlier dubbed this way of thinking about the function of names, pejoratively, the “‘Fido’-Fido Principle” (p. 70).

<sup>9</sup>Cf. Kripke (1972), in particular. For exceptions, see Evans (1973) and Katz (1994). Of course, there are many other other philosophically interesting questions regarding names as well, such as whether they are in fact devices of direct reference and how they get associated with particular individuals. As indicated in footnote 4, I remain neutral on the former here; I also remain neutral with regard to the latter.

<sup>10</sup>Cf. Bach (1991, 2002) and Katz (1994). As was noted in chapter 3, footnote 40, I am leaving aside the hard question of what is required for someone to bear a name (or, more precisely, to bear a name in context). To be clear, I take it that this is likely to have to encompass more than just having been officially baptized with the name N.

by considering which objects bear that name). In other words, despite its having been developed primarily with demonstratives in mind, the neo-Gricean theory would seem well-situated to help account for the reference of names as well. What's more, given Grice's general ambitions for his theory of meaning, it would seem odd for the neo-Gricean not to hope to extend her theory in this way. And, in fact, nothing about how Bach and others have gone about precisifying the Gricean theory so as to account for the meaning of referential terms in context appears to hinge on anything unique to demonstratives.

But perhaps this is too quick; perhaps an appeal to referential intentions isn't necessary after all. Consider that on one prominent theory of names, put forward in Kaplan (1990), names refer, at most, to just one thing. That is, according to Kaplan, there is no one name 'Jones'. Rather, there are any number of homophonic/homographic names, 'Jones<sub>1</sub>, Jones<sub>2</sub>,...', each corresponding to some person in our speech community. That person is in some sense still named 'Jones', but this is just to say that his name belongs to a homophonic/homographic equivalence class, along with all the other 'Joneses' (Kaplan 1990, p. 110-12). In other words, on Kaplan's theory, the term 'Jones' is ambiguous, just like the term 'bank'. So, just as we commonly posit there are really two distinct words, 'bank<sub>1</sub>' and 'bank<sub>2</sub>', that happen to be written and pronounced in the same manner, Kaplan posits there are any number of names 'Jones<sub>i</sub>' (what he calls 'common currency names') which also happen to be written and pronounced in the same way (Kaplan calls these ways 'generic names'). If Kaplan's theory is right, then it looks like the neo-Gricean needn't be committed to referential intentions fixing the reference of names in context; once we fix which name is uttered, we fix who is referred to.<sup>11</sup>

Would this make a difference from our point of view? No. Here's why: suppose that what uses of names require is disambiguation in context. That is, when someone uses the generic

---

<sup>11</sup>Note that Grice's actual treatment of proper names seems to be incompatible with accepting a common currency name/generic name distinction (cf. Grice 1989c, pp. 130-37). Still, there would seem to be no deep inconsistency in pairing such a distinction with the relevant aspects of Gricean theory in which we are interested here—thus making such a theoretical combination worth considering in the present context.

name ‘Jones’, what we need to know is which name ‘Jones<sub>*i*</sub>’ to associate with that use. The standard way of disambiguating uses of ambiguous terms is again via an appeal to the speaker’s intentions. And, in fact, given Grice’s general commitments about speaker meaning and occasion meaning, the neo-Gricean should be committed to this too. So, since what a speaker means on a given occasion with a use of the term *x* is supposed to be determined by her reflexive intentions to use *x* in a particular manner, then, on a Gricean-Kaplanian theory of names, which name was used will depend on precisely this sort of intention as well.

The Gricean-Kaplanian view of names will thus include the claim that the reference of a use of a given generic name is fixed by the speaker’s *disambiguating intentions*—where these are intentions to use a particular name  $N_i$  (pronounced ‘N’) and to have the listener recognize that this is the name the speaker is using, in part by recognizing the speaker’s intention to use that particular common currency name. Strictly speaking, this is enough for our purposes: the puzzle involving names introduced below can instead be run for the Gricean-Kaplanian as a puzzle about these disambiguating intentions, rather than as a puzzle about referential intentions *per se*.

But notice the following: the main criterion by which names in a public language are individuated, on a Kaplan-style picture, is *via their referent*.<sup>12</sup> One might be tempted to say ‘via the chain by which we have acquired them’, but that chain will vary, often quite markedly, from speaker to speaker. Thus, the speaker cannot have in mind that the listener come to recognize her use of a name by identifying *the same* causal chain in her own vocabulary. More plausibly, the speaker will intend for the listener to identify a certain person as the referent—and thereby to identify which name she was uttering. If that is the case, however, then there is little ground for the neo-Gricean to claim that there is any important difference between referential and disambiguating intentions, at least when it comes to names.

In this section we have thus established: first, that on any extant theory of names, one

---

<sup>12</sup>Strictly, for Kaplan (1990), names are individuated via their creation conditions (pp. 115–17). However, since (for non-empty names at least) these come bound together with a referent that the name was created to refer to, we can safely ignore this additional bit of complexity.

will need some explanation of what fixes the reference of a name in context, or else of what disambiguates that name in context; and, second, that the same sort of reasoning that led to the neo-Gricean theory of reference in the first place should push the neo-Gricean to extend her theory to include uses of names. Thus, neo-Griceans should expect for the reference of names to conform to Grice's reflexivity requirement. As it turns out, however, names can refer even when the speaker's referential or disambiguating intentions (hereafter just 'referential intentions') fail to meet this condition.

## 4.5 Sneaky Intentions

Consider the missives of the Delphic Oracle or the ways in which witnesses sometimes try to offer misleading testimony without technically perjuring themselves.<sup>13</sup> These are instances of what I call 'sneaky meaning'. Sneaky meaning, I take it, arises when a speaker intends for her listener to take her to mean one thing, while *privately committing* herself to her utterance meaning something else. Sneaky meaning thus depends on *sneaky intentions*. As it happens, some particularly clear instances of these sorts of intentions—intentions for the listener to recover  $p$  while simultaneously committing oneself to  $q$ —involve uses of referential terms like names and demonstratives.

I start with a case involving names. You come into my office and tell me that you are hungry. We often eat lunch together, but, as it happens, I've already eaten today. I tell you:

- (5) Sam is eating lunch at LuValle.

Now, suppose that there are in fact two Sams in our department, Sam<sub>1</sub> and Sam<sub>2</sub>. For whatever reason, you aren't particularly fond of Sam<sub>1</sub>, and I know this about you. In fact, it

---

<sup>13</sup>For discussion of this latter sort of case, see Solan & Tiersma (2005), Saul (2012), and Asher & Lascarides (2013). Asher and Lascarides helpfully term these cases of 'strategic conversation'. Note that there is some overlap here with the legal notion of an *adversarial* or *hostile* witness. Crucially, however, there are differences as well: witnesses need not actually be acting in a non-cooperative fashion in order to be classified by the court as adversarial. Rather, in certain circumstances, witnesses can be so classified purely in virtue of their relationship to one or more of the parties to the dispute.

is Sam<sub>1</sub> who is eating lunch at LuValle at the moment. I however (correctly) expect that you will take me to be referring to Sam<sub>2</sub>. After all, I am your friend and I don't generally provide you with useless or misleading information. Unbeknownst to you though, I've reached the end of my rope with your dislike of Sam<sub>1</sub>, which I've always found rather silly. So my hope is that you'll go down to LuValle, see Sam<sub>1</sub>, feel obliged to have lunch with him, and then realize that your dislike of him is ill-founded. Still, I'm concerned not to lie to you outright. You are my friend after all, and I make it a point not to lie to my friends—though I am not above deceiving them on occasion, at least for their own good.

Plausibly, I uttered something true in this 'Sam<sub>1</sub>-Sam<sub>2</sub>' case. While you might, for instance, try accusing me of lying, I have a rather good claim not to have been: I meant to be referring to Sam<sub>1</sub>, who after all *was* eating lunch at LuValle. What's more, I can make this claim without being disingenuous. Of course, I was trying to deceive you with my original utterance. But we can and do deceive each other without lying; just consider cases where I implicate something false while, strictly speaking, saying something true.<sup>14</sup> In the Sam<sub>1</sub>-Sam<sub>2</sub> case, I have undoubtedly tried to mislead you, but I have not lied to you. The most natural explanation for this is that I simply didn't say anything false. Nor did I take myself to be saying anything false in uttering (5).<sup>15</sup>

My intention here would appear to be a *sneaky* one: I intend to commit myself to  $o_1$

---

<sup>14</sup>To borrow an example from Adler (1997): the murderer at the door asks where Joe is and you respond with "He's been hanging out around the Nevada a lot," where the Nevada is the local diner (pp. 437–38). Joe is in fact hiding in the basement. Still, you haven't lied to the murderer (assuming that Joe has recently been spending quite a bit of time at the diner). You've just mislead her via a false conversational implicature.

<sup>15</sup>On basically any extant account of lying, saying something the content of which one believes to be true is sufficient to prevent a speaker from lying. See, for instance, Chisholm & Feehan (1977), Carson (2006), Fallis (2009, 2012), and Saul (2012). Jeff King (p. c.) has suggested that this judgment might equally well be explained by positing that the speaker hasn't said something determinate at all in this case, in virtue of her split intentions—that is, by positing that reference simply fails in cases like this one. This might well be correct, but notice that in certain contexts we *can* in fact lie by uttering sentences that include terms which fail to refer. Just consider a situation in which I tell you "I cannot make it to your party because my sister is coming into town," but where I in fact have no sister. Since this utterance is pretty clearly a lie, reference failure cannot always be exculpatory and King's suggested alternative explanation is shown to be incomplete. What's more, the suggestion that there is reference failure in (5), and thus that (5) expresses nothing determinate, just seems wrong; were Sam<sub>1</sub> to have finished lunch and left LuValle at the point when I utter (5), my utterance would be flat-out false, not truth-valueless.

being the referent while simultaneously intending for you, the listener, to identify  $o_2$  as the referent (where  $o_1 \neq o_2$ ). In other words, the speaker here intends to refer to  $\text{Sam}_1$ —and succeeds in doing so—while simultaneously understanding that she is putting the listener in a situation where he is overwhelmingly likely to recover  $\text{Sam}_2$  as the referent. What’s more, this is exactly what the speaker intends for the listener to do. This sneaky intention to refer to  $\text{Sam}_1$  thus conforms to neither Grice’s reflexive criterion for meaning nor to the neo-Gricean’s reflexive criterion for referring. It is an intention for the listener to recover  $\text{Sam}_2$  as the referent, but one which nonetheless serves to fix  $\text{Sam}_1$  as the referent.<sup>16</sup>

This leaves the neo-Gricean with a problem: the only potentially reflexive—and hence, by her lights, genuinely referential—intention in the vicinity is the intention to refer to  $\text{Sam}_2$ . It is  $\text{Sam}_2$ , after all, that the speaker intends for the listener to identify as the referent. But if anything at all is clear about this case, it is that the speaker’s use of ‘Sam’ does not refer to  $\text{Sam}_2$ . The neo-Gricean thus looks stuck predicting either that there is no genuinely reflexive intention at all to be found here, and thus that reference fails, or else that the speaker’s use of ‘Sam’ refers to  $\text{Sam}_2$ . Unfortunately, neither of these look like good options.

Interestingly, we needn’t move all the way to a case of full-fledged sneaky intentions in order to exert pressure on the neo-Gricean theory of reference. Consider, for instance, a variant of the  $\text{Sam}_1$ - $\text{Sam}_2$  case in which I don’t actually intend to deceive you about which Sam is having lunch; I’m just too lazy to clarify which Sam I intend to be the referent. I know that there is a good chance that you’ll mistakenly think that I’m talking about  $\text{Sam}_2$ . But I also believe there to be a remote chance that you passed by  $\text{Sam}_2$ ’s office today, in which

---

<sup>16</sup>Note that Grice himself used the term ‘sneaky intention’ to refer to something rather different: namely, an intention that the listener accept  $p$  on certain grounds, but that he think that he is supposed to accept  $p$  on some other grounds (Grice 1989e, p. 302; see also Bach & Harnish 1979, pp. 100–03). Crucially, here it is always the same  $p$  that the speaker is trying to convey and that the listener is supposed to accept, whereas in our cases of sneaky reference, the listener is supposed to accept  $q$  instead of  $p$ . Thus, while Strawson, Stampe, and Schiffer—to whom Grice was responding in banning ‘sneaky intentions’—were aiming to show that one could fulfill all of Grice’s conditions on meaning while not meaning  $p$ , my argument here is meant to show that one can in fact have nothing even remotely resembling a reflexive intention while *still* meaning  $p$ . In other words, these earlier examples of ‘sneaky intentions’ all targeted the sufficiency half of Grice’s definition; my examples, in contrast, target the necessity half.

case you might have noticed the note on his door saying that he is out sick. So perhaps you will realize that I'm talking about Sam<sub>1</sub> after all. Thus, while I don't take it to be *impossible* that you might figure out that I'm talking about Sam<sub>1</sub>, I don't take it to be even remotely likely that you'll figure this out—nor do I take any concrete steps to nudge you in the right direction. This intention thus violates the Gricean reflexive requirement on meaning every bit as much as the above sneaky intention does: while I may think it possible for you to recover my intended referent, it is doubtful that I actually *intend* for you to recover that referent. Rather, I am basically indifferent to your success or failure. Still, as before, I am clearly referring to Sam<sub>1</sub>. Sneaky intentions thus aren't alone in threatening the neo-Gricean theory of reference; *lazy intentions* will serve equally well.<sup>17</sup>

In chapter 2, we already saw a parallel case involving the use of a demonstrative: the Camry-McLaren case. For the sake of clarity, I will reintroduce that case here. You and I both work for a large investment bank. We've just been paid an obscenely large bonus, and we are both at the office very early one day. We run into each other near a window overlooking the parking lot, which is nearly empty at this hour. There is, however, a small cluster of cars which includes several nondescript vehicles and, in addition, a McLaren F1. You ask me what I decided to do with my bonus, and I respond by pointing out the window at the cluster of cars and saying:

(6) I bought that.

In fact, I bought the Toyota Camry parked next to the McLaren, and I intend to refer to the

---

<sup>17</sup> Neale (2004) offers a weakened version of Grice's reflexivity requirement that might seem to get the lazy Sam<sub>1</sub>-Sam<sub>2</sub> case right—and even perhaps to accommodate the sneaky version of the case. Essentially, Neale proposes that the minimum bar for intending that one's listener identify the referent in way *w*, by means of recognizing that the speaker so intends for him to act, is very low. The speaker needn't expect that the listener will succeed in this; rather, she must simply *not believe it to be impossible* for the listener to succeed in this task (Neale 2004, pp. 80–81). The lazy Sam<sub>1</sub>-Sam<sub>2</sub> case should clear this bar. And, indeed, the sneaky version might clear this bar as well; the listener, after all, might be suspicious for some reason, or might have just come from having a coffee at LuValle. The problem at this point is to say what exactly this weakened constraint would serve to rule out. For instance, Kaplan shouldn't believe it to be impossible for his listener to identify the picture of Carnap as the referent on the basis of his use of 'that' in the Carnap-Agnew case. Neale's proposed weakening thus threatens to undermine one of the main advantages of the neo-Gricean account: its ability to successfully account for the Carnap-Agnew case.

Camry, not the McLaren. That said, I've decided to have a bit of fun with you.<sup>18</sup> Even in a situation like this one, I take it, I have plausibly succeeded in referring to my Camry. But if that's right, then exhibiting the Gricean reflexive form cannot be required for an intention to fix the reference of a demonstrative in context.

Recall that a lazy version of this case could be generated as well: I'm not intending to have any fun with you here, I'm just in a hurry. Still, one of the non-McLaren's, namely my Toyota Camry, is markedly newer than the rest. What's more, I'm known to be quite thrifty, so purchasing a McLaren would be very unexpected behavior on my part. On the basis of this, I take it to be *possible* for you to work out which car I intend to refer to—I just don't take it to be very likely. Unfortunately, I lack the time to point you more squarely in the right direction. Again, I plausibly succeed in referring to my Camry in this version of the case, even though I actively doubt that you'll succeed in recovering that Camry as the referent in this context, given what I say and how I gesture.<sup>19</sup>

What the above cases illustrate is that both names and demonstratives can refer to objects even when the speaker lacks any sort of intention for the listener to identify the

---

<sup>18</sup>We can even imagine how this conversation might continue:

- i. Wow, that must have cost a pretty penny.
- ii. Yeah, I was feeling pretty flush—so I even sprung for the power windows and AC.
- iii. Wait, McLarens don't come with power windows and AC!?
- iv. How would I know? I've never bought one. I bought that sweet-looking Camry next to it.

Part of the humor of this exchange hinges precisely on the listener's misunderstanding the speaker in a particular way, a way that the speaker in fact anticipates. If the speaker's use of 'that' either really did refer to the McLaren, or simply failed to refer, it is unclear how the listener could have misunderstood the speaker in the right way—and hence why this case exhibits the particular sort of humor that it does.

<sup>19</sup>It is worth noting that these cases are equally problematic for King (2013a,b)'s 'coordination account' of demonstrative reference. Recall that, according to King, reference succeeds only when (i) the speaker intends to refer to some object *o*, and (ii) a competent, attentive, reasonable listener who is fully aware of the common ground of the conversation would in fact identify *o* as the referent. (King 2013b adds: (iii) that this listener must also identify the referent in the way that the speaker intended for him to.) The problem for King is that, in both the sneaky and lazy Camry-McLaren cases, speaker's intentions violate his condition (ii)—yet reference in these cases succeeds nonetheless. King (p. c.) disputes the claim that reference succeeds here. However, as noted in footnote 15, not only is it counterintuitive that reference fails in cases like these, reference failure would bring with it some nontrivial costs. In particular, reference failure would deprive us of the natural explanation for why I am not lying to you in the sneaky version of the Camry-McLaren case: I am saying something I believe to be true, even though I doubt that you will properly understand me.

relevant object as the referent. In other words, reference can apparently succeed even in the absence of a genuine neo-Gricean referential intention.<sup>20</sup> If this is right, it constitutes a serious threat to both the neo-Gricean theory of reference and to the Gricean theory of meaning more broadly. But perhaps there is some way for the neo-Gricean to escape this threat. In the next section, I argue that the prospects for such an escape look dim.

## 4.6 Defensive Maneuvers

In this section, I canvass four ways in which the neo-Gricean might try to sidestep the threat posed by the apparent semantic efficacy of sneaky and lazy intentions. I argue that none of these defensive maneuvers look particularly promising. For the sake of simplicity, I will limit my discussion here to just the Sam<sub>1</sub>-Sam<sub>2</sub> cases. However, analogous considerations will hold throughout for the Camry-McLaren cases.

First, the neo-Gricean might simply deny that the speaker refers to Sam<sub>1</sub> in both the sneaky and lazy versions of the Sam<sub>1</sub>-Sam<sub>2</sub> case (presumably, she would also want to deny that the speaker refers to Sam<sub>2</sub>). But the judgments in these cases seem clear enough and, in fact, mutually reinforcing: the speaker refers to Sam<sub>1</sub>, the speaker says something true of Sam<sub>1</sub>, and the speaker hasn't lied. In the absence of some further grounds for rejecting these judgments, I can see no reason to take this denial all that seriously.<sup>21</sup>

Second, the neo-Gricean might claim that these judgments rely on conflating *speaker reference* and *semantic reference*.<sup>22</sup> But according to the neo-Gricean's own theory, speaker reference and semantic reference always coincide: with the latter depending on, and ulti-

---

<sup>20</sup>For related arguments against the Gricean program—aimed at implicatures rather than at semantic contents—see Asher & Lascarides (2013).

<sup>21</sup>Further reinforcement for these judgments can be had by considering whether the speaker would have perjured herself had she uttered (5) under oath. Plausibly, she would not have perjured herself. This may not constitute genuinely independent evidence, however, since the test that judges standardly deploy to determine whether a witness is offering perjurious testimony is to consider whether or not she is lying. See, for instance, *Bronston v. United States* 409 U. S. 352 (1973). See also 18 U. S. C. §1621.

<sup>22</sup>This is basically the strategy pursued in Kripke (1977), where Kripke argues against Donnellan (1966)'s distinction between referential and attributive uses of definite descriptions.

mately identical to, the former. Thus, there is no gap to be exploited here. Rather, if the speaker's use of 'Sam' successfully referred to either Sam<sub>1</sub> or Sam<sub>2</sub>, the neo-Gricean theory dictates that the speaker herself must have referred to that very same Sam.

Third, the neo-Gricean might claim that the speaker is only *making as if to refer* to Sam<sub>1</sub>, while actually referring to Sam<sub>2</sub>. But this seems like a pure mis-diagnosis of the case: in classic cases of making as if to say—such as sarcasm and certain forms of irony—the speaker acts as though she is putting forward some content *p*, to which she explicitly intends *not* to commit herself, in order to communicate *q*. All of this, however, is supposed to be clear to the listener. In the present cases, the speaker has no expectation that the listener will construe her as referring to Sam<sub>1</sub>. Nor does she expect that by recognizing her feigned intention to refer to Sam<sub>1</sub>, the listener will somehow come to realize that she is actually referring to Sam<sub>2</sub>. This 'making as if to say' defense thus seems to be going hopelessly astray.<sup>23</sup>

Without much prospect of denying the relevant judgments, and with the two classic Gricean escape protocols mis-firing, there is still one option left open to the neo-Gricean: accept the judgments in each of the two cases but deny that they should be taken seriously in offering a theory of reference. Interestingly, Neale (2004) has already offered an argument roughly along these lines, a challenge to the advocate of a non-Gricean theory of reference to clarify what exactly her notion of 'reference' can be used to explain.

Here is Neale's challenge:

[A] distinction between what *A* referred to and what *A* *intended* to refer to is not one obviously lacking a point. So, in the first instance we should separate (i) *who or what A intended to refer to by an expression X on a given occasion*, and (ii) *who or what a rational, reasonably well-informed interpreter in B's shoes thinks A intended to refer to by X on that occasion*. In cases where (i) ≠ (ii), we *could* argue about which of (i) or (ii) or some third thing has the 'right' to be called *the person or thing referred to*, but what would be the point?...There is simply

---

<sup>23</sup>Note that the neo-Gricean cannot claim that the speaker is making as if to refer to Sam<sub>2</sub> while actually referring to Sam<sub>1</sub> (which seems plausible enough, in fact), because then Sam<sub>1</sub> would remain the referent—in violation of the neo-Gricean's requirement that reference be fixed by a genuine referential intention, one that exhibits the Gricean reflexive form.

no rôle for a transcendent notion of what was referred to upon which (i) and (ii) converge when all goes well. (Neale 2004, p. 80; italics in original)

This requires a bit of explanation. The first thing to note is that Neale is parting ways a bit from earlier neo-Griceans, such as Bach. Like Bach, Neale requires that genuine referential intentions must exhibit the Gricean reflexive form.<sup>24</sup> Unlike Bach, however, Neale imposes a further condition on successful reference: not only does it require a genuine referential intention, the intended referent must also be what a semi-idealized listener would recover, in context, as the referent.<sup>25</sup> No matter. For the sake of argument, let us simply grant Neale this neo-Gricean\* theory of reference—a slightly strengthened version of the neo-Gricean theory as introduced above. Neale’s challenge to the anti-Gricean, then, amounts to this: specify what, beyond what the neo-Gricean\* theory can already explain, one might possibly want to know about the behavior of referential terms in context. And specify further how some non-Gricean notion of reference is going to help explain that behavior.

The proper response to this challenge, it seems to me, is two-fold. First, it is worth pointing out that we might well be interested in a notion of reference that helps to predict the truth and falsity of utterances in context; in fact, many philosophers interested in the notion of linguistic reference seem to have been interested in just this. Neither of Neale’s proposed notions of reference will suffice for this task. To see why not, consider what each of Neale’s two types of reference amount to in the Sam<sub>1</sub>-Sam<sub>2</sub> case: since the speaker fails to exhibit a properly reflexive intention to refer to Sam<sub>1</sub>, Neale will claim that she doesn’t *really* intend to refer to Sam<sub>1</sub>; and a reasonable, attentive listener would presumably recover Sam<sub>2</sub> as the referent. But in order to explain why (5), i.e. “Sam is eating lunch at LuValle,” is true in the circumstances described, we need for Sam<sub>1</sub> to be the referent. Thus, there seems to be at least some reason to appeal to a ‘transcendental’ notion of reference after all.

---

<sup>24</sup>Neale (2004) imposes a reflexive condition on referential intentions earlier in his paper, at pp. 77–78. For further clarification of Neale’s understanding of Gricean reflexivity, see also Neale (1992, pp. 32–37).

<sup>25</sup>In chapter 3, I argue at length against the propriety of appealing to semi-idealized listeners in semantics. Similar considerations apply here. However, I leave those considerations aside for present purposes.

But perhaps this response has merely sidestepped the real force of Neale's challenge. After all, part of Neale's worry might be that the abstract notions of utterance-truth and -falsity that philosophers so often debate are ill-situated to play any explanatory role in a broader theory of language use. They are, we might say, mere philosopher's fictions—in which case, being able to accurately predict judgments involving these notions shouldn't amount to much of a virtue. If this is Neale's fundamental worry, then my initial response would seem to be inadequate, for it does nothing to address this version of Neale's challenge.

The problem with this second version of Neale's challenge is that it is based on a contentious, and quite possibly false, premise: that only philosophers care about the notion of utterance-truth or -falsity. In fact, there seems to be ample reason to doubt this claim. The Greeks, for instance, seem to have cared quite a bit about whether the Delphic Oracle was telling them true, if extraordinarily difficult to decipher, things—or if, on the contrary, she was simply uttering falsehoods. Likewise, we tend to care quite a lot about whether witnesses' statements under oath are true or false, not to mention whether our friends' and acquaintances' statements are (though not under oath, ordinarily). Thus, while Neale's worry might seem to have some initial bite, it is highly unclear whether his challenge can be sustained. Plausibly at least, judgments on utterance-truth and -falsity are both a widespread and important part of our overall linguistic practice, bound up with judgments on other significant matters like trustworthiness. Explaining such judgments, then, is a reasonable desideratum to which a theory of language use might aspire.

I conclude from this discussion that, at present at least, we have little reason to discount the judgments in the sneaky and lazy Sam<sub>1</sub>-Sam<sub>2</sub> cases. Rather, it seems that, *pace* the neo-Gricean theory of reference, sneaky intentions can indeed serve to fix reference.<sup>26</sup> What

---

<sup>26</sup>Another possible response on behalf of the neo-Gricean, suggested to me by Jessica Pepp, would be to treat the speaker in the Sam<sub>1</sub>-Sam<sub>2</sub> case as though she were talking to herself, rather than to the listener. This would allow her to reflexively intend to refer to Sam<sub>1</sub>, in virtue of her (trivial) intention that she herself recover Sam<sub>1</sub> as the referent. The problem with this response is that it entails giving up on the very thing that is distinctive of Gricean theories of meaning: the thought that meaning is determined by speakers' *other-directed* intentions. Besides, this description seems like a pure misdiagnosis of the case. The speaker fully intends to have a particular effect on her listener, just not the one that Griceans predict she should.

we should ask at this point is whether it is possible to offer a theory of reference capable of accounting for both the Carnap-Agnew case, as the neo-Gricean could, and cases involving sneaky and lazy intentions. I turn now to the task of demonstrating that the constraint theory can accomplish both these things.

## 4.7 The Constraint Theory Revisited

In chapter 3, I introduced the constraint theory of reference. Essentially, that theory follows Grice insofar as it takes reflexive intentions to be characteristic of cooperative communication. In contrast to the neo-Gricean theory of reference, however, the constraint theory does not treat reflexivity as a necessary condition for a speaker's intending to refer to something on a particular occasion. Rather, what the constraint theory posits is that, in order to help make their intentions known to their listeners, speakers choose specific referential terms—which, in turn, are associated with particular public commitments regarding the referent. Different sorts of referential terms are associated with different public commitments. Cooperative speakers rely on these commitments to help make their referential intentions evident; uncooperative speakers rely on these commitments to help mask their intentions. According to the constraint theory, what matters for fixing reference is not whether the speaker is aiming to be cooperative, but rather (i) that she have an intention to refer to some particular object *o*, and (ii) that *o* satisfy whatever public commitments are associated with the particular term that the speaker has chosen.

### 4.7.1 Cooperative Plans and Deceptive Plans

I have already introduced the notion of a 'communicative plan'. Now, however, we are in a position to see that these come in two flavors: *cooperative* communicative plans and *deceptive* communicative plans. In the last chapter, we dealt exclusively with cooperative plans—plans that aim at achieving transparency. In this chapter, on the other hand, we have been dealing

largely with intentionally opaque ones.

Consider, for example, speaker's plan to deceive the listener in the sneaky Sam<sub>1</sub>-Sam<sub>2</sub> case. The speaker doesn't merely intend to utter a sequence of words with a particular set of meanings attached. Nor does she intend to transmit a particular thought to the listener. Rather, her plan looks something like: utter (5) with 'Sam' referring to Sam<sub>1</sub>, and while doing nothing to dissuade the listener from taking the referent of 'Sam' to be Sam<sub>2</sub>. Her intention to deceive the listener about who is eating lunch at LuValle thus subsumes an intention to utter a particular sentence, which itself subsumes intentions to utter each word with a particular meaning or referent attached to it. It is this sort of planning structure that allows the speaker to coherently aim to prompt the listener identify Sam<sub>2</sub> as the referent, while simultaneously committing herself to Sam<sub>1</sub> in fact being the referent.

Crucially, most linguistic plans are more transparent than this one. Speakers typically say 'P' in order to communicate *p precisely because* 'P' is a standard way of conveying the thought that *p* to one's listener. In the sneaky Sam<sub>1</sub>-Sam<sub>2</sub> case, (5) is indeed a standard way of communicating that Sam<sub>1</sub> is eating lunch. The problem is that it is also a standard way of communicating that Sam<sub>2</sub> is eating lunch. While it is perhaps unclear whether it is right to say that the speaker aimed to communicate anything at all here, the following is clear enough: the speaker intended for the listener to arrive at a certain interpretation of the utterance by relying on that listener's propensity to react as though he were in a cooperative situation. In other words, the speaker intended for the listener to mistakenly treat her utterance as a cooperative one, and thus to associate her use of the name 'Sam' with the most obvious or salient candidate-Sam relative to the context. One thing this should make clear is that sneaky intentions are fundamentally parasitic on the practice of using utterances and expressions cooperatively. Understanding sneaky uses of language therefore requires first understanding how more transparent uses of language work.<sup>27</sup>

---

<sup>27</sup>See Tomasello (2008) for discussion of some related ideas about the relationship between deceptive and non-deceptive uses of language.

Thankfully, we have already made some progress on this latter question. In cases of transparent communication, speakers utter ‘P’ in order to convey one of the things that ‘P’ can standardly be used to mean. Thus, we can conceive of a speaker’s utterance of ‘P’ as a signal that the speaker intends for the listener to construe her as intending to communicate one of the standard meanings of ‘P’. In other words, if the speaker has as part of her communicative plan an intention that the listener come to believe  $p$  (or entertain  $p$ , or add  $p$  to the conversational record)<sup>28</sup>, she can signal this to the listener by saying ‘P’. When there is more than one  $p_i$  associated with ‘P’, cooperative speakers will need to take further steps to isolate this particular content as their intended meaning—that is, unless the context alone somehow serves to make it clear that  $p_i$  is what the speaker must have meant with her utterance of ‘P’. Deceptive speakers, on the other hand, exploit this slack between the signal generated by an utterance of ‘P’ and the various meanings that a listener might reasonably associate with it.<sup>29</sup>

The constraint theory thus conceives of utterances and their parts as, primarily, tools for signaling one’s communicative intentions to one’s listeners. The question now is how this signaling is reflected at the level of reference. The suggestion explored in the last chapter ran as follows: the signals generated by uses of referential terms are generated by the public commitments associated with those particular terms, via convention. These public commitments, in turn, serve to filter the speaker’s intentions for just those intentions that satisfy the relevant commitments. This means that only intentions that satisfy the relevant filter can serve to fix reference in context. With names, for instance, the use of the

---

<sup>28</sup>Reasons for adopting these alternatives can be found in Neale (1992) and Stalnaker (1978), respectively.

<sup>29</sup>All this is compatible with the claim that meanings (like  $p$ ) are conventionally associated with utterances like ‘P’, in something like the sense of Lewis (1969). That is, roughly, associations of meanings with utterances are regularities in behavior that agents conform to in response to a recurrent type of situation—in part on the basis of their expectation that other agents will do so as well, and that these agents prefer to conform too over other behavioral options, so long as each other agent also so conforms (Lewis 1969, p. 58). The point is just that, since ‘P’ may be associated with multiple  $p_i$ ’s, the speaker may need to do more than just utter ‘P’ in order to indicate which of these she means on a given occasion. Her utterance of ‘P’ will constitute a part of her signal to the listener, and will serve that function via the range of meanings conventionally associated with it. What remains, assuming that the speaker is aiming to be cooperative, is to indicate which of the meanings in this range she is trying to convey in relevant context.

name N serves to filter the speaker’s intentions for just those intentions to refer to objects that actually bear the name N.<sup>30</sup> Likewise with demonstratives: the use of a demonstrative serves to filter the speaker’s intentions for just her intentions to refer to things that are proximal or distal at the context (depending on which demonstrative was used). If that use is accompanied by a gesture, then a further filter is imposed to the effect that the object to which the speaker intends to refer must fall within the range of values indicated by her gesture. Only intentions that pass through these filters can determine the reference of a name or demonstrative in context. The use of a name or demonstrative thus refers to some object  $o_1$  only when (i) at least one intention to refer to  $o_1$  passes the filters associated with the particular name or demonstrative used and (ii) no intention to refer to  $o_2$  (where  $o_1 \neq o_2$ ) also passes those filters. If there is no  $o_i$  that satisfies both (i) and (ii), then reference fails. This, then, is the sense in which public commitments serve to constrain how intentions fix reference in context.<sup>31</sup>

#### 4.7.2 Sneaky Reference and Lazy Reference

Now it is time to show how the constraint theory can handle the various cases introduced above. Of particular interest will be to illustrate that it is precisely the ways in which the constraint theory departs from the neo-Gricean theory are what allows it to make better

---

<sup>30</sup>Were we to understand this naming condition very strictly—such that in order to bear the name N, one must actually be named N—then one might rightly view the constraint theoretic approach to names as a modification of the sort of nominal descriptivism outlined in Russell (1910). There, Russell suggests that ‘Julius Caesar’ means ‘the man whose name was Julius Caesar’ (p. 119). The idea would be that, once we realize that there can be multiple people named N, we can no longer endorse a meaning claim quite like Russell’s. However, we might still think that names restrict the range of individuals to whom one can refer to just those who have been properly so-named. In contrast to this sort of neo-Russellian, I suspect that being named N is only *one way* (though perhaps the paradigm way) of bearing the name N in context; other possibilities might include having N as a nickname or being called N for the purposes of the conversation. For discussion, see Matushansky (2006a,b), Cumming (2007), Fara (2013a), and Jeshion (2013a,b).

<sup>31</sup>Note that this notion of a public commitment is aimed at capturing a certain normative status that speakers take on by using certain terms; the notion is not meant to describe a mental state or attitude that speakers might or might not be in while using such terms. Cumming (2013) invokes a similar notion to explain the differences that arise in the truth-conditional contents associated with uses of definite and indefinite descriptions—in spite of both being typically used to convey thoughts about particular objects. Definite descriptions, the idea goes, publicly commit the speaker to coordinating with the listener on a particular object as the referent, whereas indefinite descriptions come with no such public commitment.

predictions in these cases.

I begin with the Sam<sub>1</sub>-Sam<sub>2</sub> case: in the sneaky version of this case, the speaker intends for Sam<sub>1</sub>, and only Sam<sub>1</sub>, to be the value of her use of the name ‘Sam’. Since this intention to refer to Sam<sub>1</sub> satisfies the filter associated with the use of the name ‘Sam’ (i.e. the intended referent actually bears the name ‘Sam’), the constraint theory predicts that Sam<sub>1</sub> is the referent of this use of ‘Sam’. Similar considerations hold for the lazy version of the Sam<sub>1</sub>-Sam<sub>2</sub> case. The constraint theory thus proves capable of handling both versions of this case with relative ease. In contrast, the neo-Gricean was forced to predict that, since the speaker lacks a properly referential intention directed at Sam<sub>1</sub>, the speaker’s use of ‘Sam’ fails to refer to Sam<sub>1</sub>.

A brief aside: it is worth emphasizing that, according to the constraint theory, a speaker can use a name like ‘Sam’ to refer to someone named Sam even when she knows that the listener is unacquainted with this person and stands no chance of recovering that person as the referent. Such uses of names are, of course, usually highly uncooperative. But, according to the constraint theory, these uses still result in successful reference. Put this way, some may find this to be an unappealing aspect of the constraint theory. In contrast, I would contend that this feature of the view is simply a reflection of how names actually work.<sup>32</sup>

Returning now to the Camry-Mclaren case: in the sneaky version of that case, the speaker intends for her use of ‘that’ to refer to her Camry. Since that Camry is relatively distal, and since it is in the general direction—and thus falls within the range of objects—indicated by the speaker, the speaker’s intention to refer to her Camry passes through the filter imposed by her use of ‘that’, combined with her gesture. The speaker’s use of ‘that’ is thus predicted

---

<sup>32</sup>Consider, for instance, introductory uses of names: if uses of names fail to refer when we know that our listeners will not be able to identify the referent, then when I introduce a name for an absent object, that use of the name will almost always fail to refer. Of course, one might try to avoid this conclusion either by treating introductory uses as somehow special, or by allowing the listener to recover the referent under the description ‘whoever the speaker had in mind’. Something like this latter strategy is in fact suggested, in a different context, in Borg (2009). Unfortunately, this strategy is also at serious risk of being self-undermining. The problem is that listeners should *always* be in a position to recover an intended referent under such a minimal description. Thus, adopting this strategy would mean giving up on the earlier explanation of the Carnap-Agnew case in terms of reflexive intentions.

to succeed in referring to her Camry, assuming that she has no competing, non-equivalent referential intentions. Similar considerations hold relative to the lazy version of the case. Once again, these seem to be the correct predictions. And, once again, the neo-Gricean was unable to deliver these predictions, since the speaker lacks a reflexive intention to refer to her Camry in either of these cases.

Finally, it is worth recalling that the constraint theory was capable of explaining the Carnap-Agnew case as well. Thus, adopting the constraint theory does not require giving up on an adequate explanation of that case. While Kaplan plausibly intends for his use of ‘that’ to refer *both* to his picture of Carnap *and* to the picture that is actually behind him, only his intention to refer to the picture behind him succeeds in fixing reference here. According to the constraint theory, this is because Kaplan’s use of ‘that’, combined with his gesture, serves to filter out his intention to refer to the picture of Carnap—leaving only his intention to refer to the picture behind him, the picture of Agnew, to fix reference.

The constraint theory thus delivers the desired results in the full range of cases we have been investigating in the course of this chapter. Not only that, but the theory was able to deliver these results while preserving a broadly Gricean picture of communication. What proved necessary was to re-conceive of one common element of communicative intentions—namely, their potentially reflexive structure—not as a requirement for referential success, but rather as a mark of cooperativeness on the part of the speaker. What referential success requires is not that the speaker aim at being cooperative, but rather that she *privately commit* herself to one or another way of understanding a referential term as the privileged interpretation. So long as this privileged interpretation falls within the looser bounds set by the speaker’s public commitments, the speaker is free to choose whatever object she likes as the referent.

## 4.8 Conclusion

The primary aim of this chapter was to show that even the basic kernel of the Gricean theory of meaning—the claim that meaning depends on reflexive intentions—is untenable. To accomplish that, I first showed how the Gricean theory of meaning applies to the theory of reference in particular. Then, I offered several counterexamples to the Gricean theory as applied to referential terms—that is, to the neo-Gricean theory of reference. Finally, I showed how the constraint theory is capable of retaining many of the virtues of the neo-Gricean theory of reference while jettisoning the most conspicuous of its drawbacks. Essentially, what the constraint theory proposes is that the conventional aspects of referential terms—conceived of here as public commitments regarding the properties that the referent must bear—limit the ways in which speakers can use these terms to refer in context.<sup>33</sup> By way of conclusion, I will briefly demonstrate that the phenomenon of sneaky meaning is not unique to names and simple demonstratives; rather, it can be replicated with complex demonstratives, ambiguous terms, and even metaphors. This observation serves to significantly deepen the challenge facing both the Gricean theory of meaning and its kin.

First, consider complex demonstratives. Suppose that there is a man in our shared visual field who appears to be drinking a martini. I somehow know that he is drinking water out of a martini glass, but that another man behind him is in fact drinking a martini out of a flask. Gesturing only vaguely in the direction of both, I tell you:

- (7) That man drinking the martini is a jerk.<sup>34</sup>

I intend for the man drinking from the flask to serve as the referent of my use of ‘that man

---

<sup>33</sup>Seen in this way, the constraint theory exhibits significant similarities to Searle (1969)’s claim that meaning is not just a matter of intention, but of intention as constrained by convention (p. 45). The difference lies in the fact that, for Searle, it is still a reflexive intention of a certain sort that serves to fix meaning—a reflexive intention aimed at eliciting a particular interpretation of the speaker’s utterance, rather than a particular belief (Searle 1969, p. 47). The constraint theory, of course, rejects this claim.

<sup>34</sup>This case is, quite obviously, a modified version of Donnellan (1966)’s classic ‘man drinking the martini’ example. Analogous cases can in fact be generated using definite descriptions, but what we should conclude from this is less clear. In particular, things are complicated by the fact that the semantics of definite descriptions—and, in particular, whether they can be used to refer semantically—is still widely disputed.

drinking the martini' here. However, I am fully aware that you are overwhelmingly likely to construe me as referring to the man drinking water from the martini glass. In fact, that's precisely what I intend for you to do. Still, it seems that my use of 'that' refers to the man drinking from the flask, and thus that I succeed in saying something true. From this, I think we should conclude that it is not just simple referential expressions (like names and the simple demonstratives 'this' and 'that') that can tolerate sneaky reference.

Analogous cases can also be generated using ambiguous terms and phrases. Suppose that you ask me what Peter is doing today, and I respond with:

(8) Peter is going for a picnic on the bank.<sup>35</sup>

I may well expect for the sentential context here to privilege the disambiguation of 'bank' as 'the side of a river', particularly if there is a river bank in the vicinity where people often go for picnics. Nonetheless, if Peter has hatched a plan to scale the local financial institution and picnic in their rooftop garden (it's a very environmentally-friendly bank, as it happens), it seems that my utterance of (8) will be true just in case I intend for my use of 'bank' to mean 'financial institution'. In other words, disambiguating intentions appear to tolerate sneakiness in much the same way that referential intentions do.

Finally, consider a case involving metaphor. Suppose that Jack, a matchmaker, is trying to convince Bill to go on a date with Stella. He tells Bill:

(9) Stella is a long-stemmed rose.<sup>36</sup>

Bill takes Jack to mean that Stella is tall and elegant, just as Jack intends for him to. In fact, Stella is neither of these things. However, Stella *is* prickly and wears too much perfume. Not only does Jack know this about Stella, this is in fact exactly what he means to commit himself to with his utterance of (9). Once more, it seems that Jack will not have lied to

---

<sup>35</sup>This example is adapted from Davis (1998, p. 46), where it is put to a rather different use.

<sup>36</sup>This example is adapted from Camp (2007, p. 197). Again, Camp puts the case to a rather different use.

Bill—in spite of Jack’s clear intention to deceive Bill with regard to what he means.<sup>37</sup>

What these examples serve to indicate is just how widespread the phenomenon of sneaky meaning turns out to be. These are not isolated cases, involving just a couple of specific sorts of referential terms—and thus the sort of thing that the Gricean might reasonably hope to quarantine. Rather, sneaky cases can be replicated not just with a variety of other referential terms, but also with linguistic phenomena, like ambiguity and metaphor, that have little or nothing to do with reference. This constitutes a serious problem for the Gricean, since she is bound to make incorrect predictions down the line for this whole range of cases. In particular, since Griceans require that utterances be accompanied by reflexive intentions in order to count as meaningful, and since the speakers in (7)–(9) exhibit no such intentions, the Gricean looks forced to conclude that all of these utterances are meaningless.

I close, therefore, with a hypothesis: it is not just reference that depends on speakers’ private commitments, but rather the occasion meaning of terms and utterances more generally. If this hypothesis is right, it serves to reinforce the fairly radical conclusion we already arrived at at the end of chapter 3: what utterances mean on particular occasions is not the right sort of thing to play a fundamental role in explaining communication. That’s because such meaning will obtain even in instances where communication should be impossible, since the speaker is intentionally failing to make the meaning of her utterance evident. What then is this sort of meaning good for? Plausibly, it helps to explain when our utterances are true or false. More speculatively, it looks like these observations serve to reinforce the case for the suggestion I offered above, to the effect that occasion meaning helps to explain the sorts of *responsibility* that speakers bear for the utterances they make. Certain aspects of that responsibility undoubtedly have to do with the attitudes that speakers bear towards their listeners. Other aspects, however, may have little to do with listener-directed attitudes. The lesson of sneaky intentions, I take it, is that speakers can commit themselves by means of

---

<sup>37</sup>Note that we probably *can* lie with metaphors in the right circumstances. If, for instance, Jack has in mind no way in which Stella might even remotely resemble a long-stemmed rose, then presumably he would be lying to Bill in uttering (9).

their utterances to the world being a certain way, even when they fully intend to leave their listeners in the dark about the contents of those commitments.

# Chapter 5

## Conclusion

*He did not want to compose another Don Quixote—which would be easy—but the Don Quixote.  
Jorge Luis Borges, Pierre Menard, Author of Don Quixote—1939*

### 5.1 Summary

In chapters 2, 3, and 4, I argued for a particular version of intentionalism about reference, the constraint theory. That argument came in two main parts: first, I explained the reasons why we should reject the primary alternatives, demonstrationism and salientism. Second, I demonstrated the advantages of the constraint theory as against other sorts of intentionalism, both in terms of its ability to account for cases of confusion and in terms of its ability to tolerate sneaky intentions. On the basis of these arguments, I concluded that we have good reason to adopt the constraint theory: the alternatives to intentionalism look weak at best, and the constraint theory represents a significant improvement over earlier versions of the intentionalist position.

Investigating the twin phenomena of confusion and sneakiness led me to further postulate that the meanings of utterances—that is, their semantic content—ought to play a rather different role in theories of language use than has typically been supposed. In particular, if the considerations attested to above are correct, then a theory of semantic content would seem

to be the wrong sort of thing to explain either communication or communicative potential, at least not directly. As we have seen, speakers can plausibly succeed in communicating some content to their listeners even where that content fails to match the meaning of the speaker's utterance, in context—as in cases of mirroring confusion. Likewise, semantic content can arise even in cases where communication is plausibly impossible, due to the speaker's deliberate sneakiness. In such cases, even if the listener does somehow manage to recover the content to which the speaker is committed, it is highly unclear that we should want to call this successful communication. Rather, it seems that the listener will have in some sense *defeated* the speaker's deceptive linguistic plan.

On the basis of such cases, I suggested that what occasion meaning is good for—beyond pairing truth-conditions with utterances—is tracking a certain sort of responsibility that speakers take on in virtue of saying the things they do. Above, I called this a *core* type of the responsibility we bear for our utterances. This was meant to put this sort of responsibility in contrast to the sort of responsibility we might bear for the various inferences that we typically license our interlocutors to draw on the basis of our utterances. One hard question in the vicinity is how to think about the *entailments* of what our utterances mean on particular occasions. My suggestion is that we should think of these as another core type of responsibility we bear for the utterances we make. Like literal meaning, these are usually difficult to deny without contradicting ourselves—or at least without implicitly retracting our previous statement. That said, non-obvious entailments are not so difficult to deny. For example, if I happen not to know that Mark Twain is Samuel Clemens, I might attribute P to the former while denying that P held of the latter. I am wrong here, but plausibly not in a way that involves flouting my basic linguistic responsibilities. This, to my mind, should be enough to motivate positing a distinction between entailments and literal meaning. However, advocates of coarser-grained notions of semantic content may want to resist at this point. I will not attempt to offer a more extensive argument against that inclination here.

Rather, limitations of both space and thought militate against extensive inquiry into the

precise nature of this sort of responsibility. I conclude what I have to say about it here with a few brief suggestions: first, it strikes me as plausible that if we bear this sort of core responsibility for the semantic content of the things we say, then this content should be the right sort of thing to figure into at least a partial explanation of when we lie to each other. That is, while perhaps we can lie via content to which we bear only a looser sort of responsibility, semantic content should figure at the core of any productive explanation of lying. Paradigm instances of lying, I would suggest, turn on semantic content, on what we literally put forward by means of our utterances. Second, this sort of responsibility might well help us to better understand the epistemology of testimony. Warrant-transmission might conceivably require not just transmission of a thought, but of a thought that the speaker is committed to in the right way—and partly on the basis of that very commitment. Further investigation of these phenomena, it seems to me, should help to reveal the ways in which meaning is grounded in very same the social phenomena that it helps to underwrite. In other words, if, as seems plausible, the constraints on what speakers can use utterances to mean are grounded in how they can use those same utterances to tell the truth, to lie, and to transmit their warrant to each other, then by better understanding the relationship between these things, we might thereby hope to come to better understand what it means for meaning to be grounded in these sorts of phenomena.

Finally, there is one last last theme running through chapters 3 and 4 that we are now in a position to make fully explicit. *Speaker reference* has often been treated, implicitly at least, as both more basic and easier to identify in context than *semantic reference* (cf. Schiffer 1981, Bach 1987, 1992b,a, 2005, 2006, Neale 1992, 2004, Heck 2013). We now have ample reason to doubt this. In fact, speakers often have multiple, conflicting intentions to refer—as we saw in both the Carnap-Agnew and Smith-Jones cases. In these cases, it makes little sense to talk about *the* speaker referent. Nonetheless, the semantic referent was clear enough in both these cases. This is already enough to suggest that analyses of semantic meaning in terms of speaker meaning are bound to fail. That suggestion was then reinforced

by considering sneaky cases, where the speaker has a univocal intention to mean something with her use of a term, but not the thing that Griceans and neo-Griceans alike predict that she should mean. I take myself, therefore, to have made at least a *prima facie* case for the following two claims: (i) metaphysically, semantic meaning is at least as basic as speaker meaning, and (ii) methodologically, investigations of semantic meaning are independent of investigations of speaker meaning. Speaker meaning, at least in the way that Griceans have wanted to conceive of it, is a fragile concept—well defined in cooperative situations involving non-confused speakers, and perhaps not otherwise. It is therefore unclear just how helpful this notion will prove to be to the project of explicating language use.

By way of conclusion, I wish, first, to briefly consider two loose ends. Then I will take stock of where things stand—and, in particular, what is left to be done if we are to fully explicate the nature of linguistic reference.

## 5.2 Why I Still Can't be Napoleon, Despite Trying

The first loose end has to do with indexicals. While it is fairly easy to see how to extend the present theory to account for uses of deictic pronouns like 'he', 'she', and 'it', so-called 'true' indexicals like 'I', 'here', and 'now' look to be more difficult. Pronouns plausibly function more or less like demonstratives—except that instead of a constraint regarding distance, these come with differing constraints on grammatical gender.<sup>1</sup> In contrast, the worry regarding true indexicals runs as follows: we have good reason to think that what these terms refer to in context does not, in fact, depend on the speaker's intentions. But the constraint theory predicts that these terms *should* be sensitive to the speaker's intentions. If this is right,

---

<sup>1</sup>Crucially, it is grammatical gender—not actual gender, let alone actual sex—that is at issue here, at least if we take seriously the fact that we sometimes use gendered pronouns in English to refer to inanimate objects (ships, for instance). Presumably, such objects lack either a real sex or a gender, and yet reference to them via a gendered pronoun can be perfectly felicitous. In other languages, the bounds of grammatical gender vary widely from those of English. Some languages, like French, lack a neuter pronoun (akin to the English 'it'), whereas others languages do possess a neuter pronoun, but still assign a gender to a wide range of inanimate objects (like German).

then it looks like the constraint theory’s claim to offer a generalizable account of referential terms has just been stopped in its tracks. I will briefly endeavor to show that this is not the case; in fact, the relevant data fails to tell between non-intentionalist accounts of the true indexicals and *highly constrained* intentionalist ones.

What are the relevant data here? I confine myself for the time-being to consideration of just the term ‘I’. As Perry (1977) observes, the fact that the speaker might confusedly believe herself to be someone else has no bearing on the truth or falsity of an utterances she might make involving the term ‘I’. For instance, suppose that a speaker who mistakenly believes herself to be Napoleon utters (1):

(1) I am Napoleon.<sup>2</sup>

Not being Napoleon, the speaker seems to have said something false; she is simply in no position to succeed in referring to Napoleon by using the indexical ‘I’. This sort of case has often been taken to represent a decisive objection to theories that bind the reference of uses of indexicals to speakers’ intentions in uttering them.<sup>3</sup> The obvious alternative is to adopt a theory along the lines proposed in Kaplan (1989b), according to which uses of ‘I’ refer to the agent at the context—that is, standardly at least, to the speaker—regardless of how that speaker intends for her use of the term to refer.<sup>4</sup> In other words, on this alternative, ‘I’ and other true indexicals are intention-insensitive.

But notice that the sorts of views that this kind of case clearly undercuts are really just *unconstrained* versions of intentionalism. The problem that (1) presents is, in effect, yet another instance of the Humpty Dumpty Problem. More specifically, this is a case of

---

<sup>2</sup>Perry (1977)’s original example was “I wrote the *Treatise*,” as said by a speaker who takes herself to be David Hume (p. 487).

<sup>3</sup>The most sustained, and forthright, advocate for this sort of view has been Predelli (1996, 1998a,b, 2002, 2005). For criticism, see Corazza et al. (2002), Romdenh-Romluc (2006), and Cohen (2013). Predelli has responded to some of his critics by claiming that they are confusing semantic reference with some variety of pragmatic content (Predelli 2002, p. 315). For reasons elaborated on in Cohen & Michaelson (2013) and Michaelson (2013), I find this response unconvincing.

<sup>4</sup>See Michaelson (2013) for discussion of why, on a Kaplan-style semantics, it prove best to conceive of the conditions of agency as shifting in certain circumstances.

referential confusion: the speaker intends both to refer to herself and to refer to Napoleon (who she takes herself to be). And these intentions point to different objects being the referent of the speaker's use of 'I'.

At this point, it should be fairly predictable how the constraint theorist is going to propose to handle this case: she is going to impose certain restrictions on how speakers can use the term 'I' to refer in context. A rough and ready version of this might run something like: uses of 'I' commit the speaker to the referent being the speaker.<sup>5</sup> So long as we suppose that, among the speaker's myriad intentions to refer when she uses the term 'I' in uttering (1) is an intention to refer to herself, the constraint theory will deliver the intuitively correct result with regard to that utterance. Specifically, the constraint theory will predict that the speaker's use of 'I' refers to her, and does not refer to Napoleon. That's because her intention to refer to Napoleon is filtered out by the constraints imposed by her use of the term 'I'.<sup>6</sup>

We are left, therefore, with no clear evidence against a constraint-theoretic approach to indexicals. Cases like (1) are compatible with non-intentionalist theories of indexicals, like Kaplan (1989b)'s. But these cases are also compatible with suitably restricted versions of intentionalism, like the constraint theory. Such cases thus offer no decisive reason for thinking that the expansionist ambitions of the constraint theory are bound to fail.<sup>7</sup>

---

<sup>5</sup> Plausibly, the rules for 'here' and 'now' will have to be slightly more complicated. One reason for this is that speakers seem to have some leeway in setting the bounds of the relevant place or duration of time. Recognizing this, one suggestion regarding the constraints associated with each of these terms would be that, while the speaker's intentions determine the exact physical and temporal bounds of the regions of space and time to which particular uses of 'here' and 'now' refer, the specific place and time of the utterance must be contained within the relevant region. 'Here', for instance, might be used to refer to a city as opposed to the particular spatial region occupied by the speaker at the time of her utterance—so long as the speaker is located within the bounds of that city at the time of her utterance. Likewise, 'now' might be used to refer to a period of several minutes rather than to the precise time of the utterance—so long as the time of the utterance falls within that larger stretch of time. Thanks to David Kaplan and Alexandru Radulescu for helpful discussion on this point. For further discussion, see Radulescu (2012).

<sup>6</sup>In fact, the constraint theory might seem to offer an advantage here over non-intentionalist theories of indexicals: it can explain why, if the speaker really has *no* intention to refer to herself with a use of 'I', she may in fact fail to refer to anything at all. As it turns out, however, non-intentionalist theories are also capable of making this prediction, though via a slightly different route: a speaker with no intention to refer to herself when using 'I' manifests a particular sort of incompetence at using that term. Thus, she may well fail to refer for reasons having to do with her lack of competence, rather than the intention-sensitivity of 'I'.

<sup>7</sup>This, of course, is not all that there is to say about the meaning of a term like 'I'. In particular, it is

### 5.3 What Grounds Meaning?

The second loose end has to do with the grounds of meaning. In chapter 4, I noted that deceptive uses of referential terms are plausibly parasitic on more standard, transparent uses of those terms. I have also suggested in several places that the standing meanings of referential terms—that is, in my terminology, the public commitments associated with those terms—are grounded not just in their cooperative uses, but also in deceptive uses. These commitments might seem to be *prima facie* incompatible. How, after all, can we obtain priority in this one sense while retaining parity in the other?

I contend, of course, is that these two claims are not at all incompatible. The first claim is about the asymmetric dependence of one sort of use on another sort of use. The second claim is about what sustains particular linguistic conventions in particular communities, or what makes it the case that, in a given speech situation, particular terms can be used in certain ways but not in others.

To clarify the first claim: sneaky uses of language like those canvassed above cannot exist without transparent uses. The reason why is that sneaky speakers are implicitly relying on their listeners to act as though they are in a cooperative scenario. If the listener does not act as though he is in a cooperative scenario, then he is unlikely to act as the speaker intended for him to. If there were no such thing as a cooperative scenario—that is, if we never reasoned about the speaker’s intended referent on the basis of the assumption that she is trying to help us identify that referent—then it would be impossible for listeners to ‘fall for’ any of the sneaky utterances introduced above. But then the speaker would have no incentive to try to use language in this way. In fact, it’s not clear that she could even intend to do so; unless

---

not clear how the present proposal might handle inscriptions and other sorts of recordings of ‘I’—where, intuitively, there is no speaker present in the context of interpretation or playback. I will not attempt to offer any sort of detailed theory of such uses of indexicals here. In Michaelson (2013), I attempted to offer a non-intentionalist theory of indexicals that adequately captures the truth-conditions of at least the core cases of recorded uses of indexicals. A constraint theoretic analogue of this theory should, in fact, prove fairly easy to generate. The most difficult cases will be those that involve *sampling* a recording and then reusing later it for a different purpose. Part of the difficulty here, however, is really that of deciding whether such cases should be accounted for by a semantic theory, or, rather, by a theory of pragmatics.

the speaker were mistaken about the situation she is in, she would know that her attempt to deceive the listener in this way is bound to fail.

The second claim, on the other hand, is about what makes it the case that one particular way of using a referential term as opposed to another is in effect in a given linguistic community at a given time. Sneaky uses of referential terms, the claim runs, can help to explain why one practice as opposed to another is in effect—just as non-sneaky uses can. In cooperative cases, the speaker is careful to work within the bounds of conventional use, so as to make her referential intentions evident to the listener. In sneaky cases, the speaker is similarly careful to work within the bounds of conventional use. Here, however, her goal is to make her use of the term, and ultimately the utterance, comport with the basic commitments these expressions require so as to preserve her claim to being both a competent speaker and a truth-teller—though perhaps not a very helpful one. Sneaky uses thus look to be linguistic acts shaped by and subject to the same sorts of conventions that cooperative uses are. If, as is fairly standard, we take it that linguistic conventions are grounded at least in part by uses of language governed by those conventions, then it looks like both sneaky and cooperative uses of language will be the right sorts of things to help ground linguistic conventions.

Put slightly differently, if what we are interested in is the question of whether we can have linguistic practices comprised entirely of sneaky uses of language, then the answer is ‘no’. Cooperative uses of language are required for sneaky uses to arise. However, if what we are interested in is whether both cooperative and sneaky uses of language can help to sustain linguistic conventions once those conventions are already in effect, then the answer is ‘yes’. Both sorts of uses are equally reliant on the existence of those conventions. Thus, to the extent that linguistic conventions are sustained by their being put to use by actual speakers, cooperative and sneaky uses of language should in fact be on par with regard to their ability to play this sustaining role.

## 5.4 Closing Thoughts

Some progress has been made. Much remains to be done. I hope to have made some inroads here into the nature of linguistic reference and the particular ways in which it depends on mental reference. Towards that end, I have provided what I hope will prove to be a productive framework for thinking about the way in which intentions determine linguistic reference, within limits. In some sense, the theory I have sketched is rather thin: I have said little, for instance, about the particulars of what it is to ‘bear a name’ or to ‘count as distal in a context’. The criticism is apt. I have had very little to say about these topics. Rather, in applying the constraint theory above, I relied on an intuitive feel for both these notions, not a worked-out definition. What this means is that, as it stands, constraint theory is really more of a theory schema than it is a fully precise, predictive theory. It is poised to become such a theory only once we start inputting these parameters in sufficient detail.

That said, the aim of this dissertation has not been to get all the details right in this regard. Rather, what I have aimed to accomplish here is to provide a framework within which these questions—what is it to bear a name, what it is to count as distal or proximal in a context—can be productively asked and, hopefully, eventually answered. The theory has essentially treated these notions as placeholders, to be filled in via careful empirical research. This puts the constraint theory in stark contrast to many of its predecessors, which tried to derive the bounds of reasonable usage from general principles governing human intentionality or human comprehensibility. In certain respects, the theory on offer here is thus far less ambitious than a number of its predecessors. It is this same lack of ambition, however, that offers the slightest sliver of hope that the present theory might, with sufficient further clarification, actually turn out to be correct.

# Bibliography

- Adler, Jonathan. 1997. Lying, deceiving, or falsely implicating. *Journal of Philosophy* 94(9). 435–452.
- Armstrong, David M. 1971. Meaning and communication. *Philosophical Review* 80. 427–447.
- Asher, Nicholas & Alex Lascarides. 2013. Strategic conversation. Forthcoming in *Semantics and Pragmatics*.
- Augustine. 1952. Lying. In Roy Deferrari (ed.), *Treatises on various subjects*, 53–110. New York: Fathers of the Church.
- Bach, Kent. 1987. *Thought and reference*. Oxford: Clarendon Press.
- Bach, Kent. 1991. What's in a name? *Australasian Journal of Philosophy* 59. 371–386.
- Bach, Kent. 1992a. Intentions and demonstrations. *Analysis* 52. 140–146.
- Bach, Kent. 1992b. Paving the road to reference. *Philosophical Studies* 67. 295–300.
- Bach, Kent. 2002. Giorgione was so-called because of his name. *Philosophical Perspectives* 16: *Language and Mind* 73–103.
- Bach, Kent. 2005. Context *ex Machina*. In Zoltán Szabó (ed.), *Semantics versus pragmatics*, 15–44. Oxford: Oxford University Press.
- Bach, Kent. 2006. What does it take to refer? In Ernest Lepore & Barry Smith (eds.), *The Oxford handbook of philosophy of language*, 516–554. Oxford University Press.

- Bach, Kent & Robert M. Harnish. 1979. *Linguistic communication and speech acts*. Cambridge: MIT Press.
- Bennett, Jonathan. 1976. *Linguistic behaviour*. Cambridge: Cambridge University Press.
- Bertolet, Rod. 1980. Demonstrations and intentions. *Philosophical Studies* 38. 75–78.
- Borg, Emma. 2009. The place of referential intentions in linguistic content. *Manuscrito* 32. 85–122.
- Bratman, Michael. 1987. *Intention, plans, and practical reason*. Cambridge: Harvard University Press.
- Burge, Tyler. 1973. Reference and proper names. *Journal of Philosophy* 70(14). 425–439.
- Burge, Tyler. 1974. Demonstrative constructions, reference, and truth. *Journal of Philosophy* 71(7). 205–223.
- Camp, Elisabeth. 2007. Prudent semantics meets wanton speech act pluralism. In G. Preyer & G. Peter (eds.), *New essays on semantics and pragmatics*, 194–213. Oxford: Oxford University Press.
- Camp, Joseph. 2002. *Confusion*. Cambridge: Harvard University Press.
- Cappelen, Herman & Ernest Lepore. 2005. *Insensitive semantics: A defense of semantic minimalism and speech act pluralism*. Oxford: Blackwell.
- Carson, Thomas L. 2006. The definition of lying. *Noûs* 40. 284–306.
- Chisholm, Roderick M. & Thomas D. Feehan. 1977. The intent to deceive. *Journal of Philosophy* 74(3). 143–159.
- Clark, Herbert H., Robert Schreuder & Samuel Buttrick. 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22. 245–258.

- Clark, Herbert H. & Deanna Wilkes-Gibbs. 1992. Referring as a collaborative process. In *Arenas of language use*, 107–143. Chicago: University of Chicago Press.
- Cohen, Jonathan. 2013. Indexicality and the puzzle of the answering machine. *Journal of Philosophy* 110(1). 5–32.
- Cohen, Jonathan & Eliot Michaelson. 2013. Indexicality and the answering machine paradox. *Philosophy Compass* 8(6). 580–592.
- Corazza, Eros, William Fish & Jonathan Gorvett. 2002. Who is I? *Philosophical Studies* 107. 1–21.
- Cumming, Samuel. 2007. *Proper nouns*. New Brunswick: Rutgers University dissertation.
- Cumming, Samuel. 2013. The semantic dilemma of indefinites. Manuscript.
- Davies, Martin. 1982. Individuation and the semantics of demonstratives. *Journal of Philosophical Logic* 11. 287–310.
- Davis, Wayne A. 1998. *Implicature: Intention, convention, and principle in the failure of Gricean theory*. Cambridge: Cambridge University Press.
- Donnellan, Keith. 1966. Reference and definite descriptions. *Philosophical Review* 75. 281–304.
- Donnellan, Keith. 1968. Putting humpty dumpty together again. *Philosophical Review* 77(2). 203–215.
- Donnellan, Keith. 1970. Proper names and identifying descriptions. *Synthese* 21. 335–358.
- Dummett, Michael. 1976. What is a theory of meaning? (ii). In Gareth Evans & John McDowell (eds.), *Truth and meaning*, 67–137. Oxford University Press.
- Elbourne, Paul. 2005. *Situations and individuals*. Cambridge: MIT Press.

- Enfield, Nick J. 2001. ‘Lip-pointing’: A discussion of form and function with reference to data from laos. *Gesture* 1(2). 185–212.
- Enfield, Nick J. 2003. Demonstratives in space and interaction: Data from lao speakers and implications for semantic analysis. *Language* 79(1). 82–117.
- Evans, Gareth. 1973. The causal theory of names. *Proceedings of the Aristotelian Society, Supplementary Volumes* 47. 187–225.
- Evans, Gareth. 1982. *Varieties of reference*. Oxford: Clarendon Press.
- Fallis, Don. 2009. What is lying? *Journal of Philosophy* 106(1). 29–56.
- Fallis, Don. 2012. Lying as a violation of grice’s maxim of quality. *Dialectica* 66(4). 563–581.
- Fara, Delia G. 2011. You can call me ‘stupid’, . . . just don’t call me stupid. *Analysis* 71(3). 492–501.
- Fara, Delia G. 2013a. “literal” uses of proper names. Forthcoming in Andrea Bianchi, ed. *New Essays on Reference*.
- Fara, Delia G. 2013b. Names as predicates. Manuscript.
- Feldman, Harry. 1986. *A grammar of awtuw*. Canberra: Australian National University dissertation.
- Field, Hartrey. 1973. Theory change and the indeterminacy of reference. *Journal of Philosophy* 70. 462–481.
- Fodor, Jerry & Ernie Lepore. 2004. Out of context. *Proceedings and Addresses of the American Philosophical Association* 78(2). 77–94.
- Gauker, Cristopher. 2008. Zero tolerance for pragmatics. *Synthese* 165. 359–371.

- Geurts, Bart. 1997. Good news for the descriptions theory of names. *Journal of Semantics* 14(4). 319–348.
- Geurts, Bart. 1999. *Presuppositions and pronouns*. Oxford: Elsevier.
- Grice, H. Paul. 1989a. Logic and conversation. In *Studies in the way of words*, 22–40. Cambridge: Harvard University Press.
- Grice, H. Paul. 1989b. Utterer’s meaning and intentions. In *Studies in the way of words*, 86–116. Cambridge: Harvard University Press.
- Grice, H. Paul. 1989c. Utterer’s meaning, sentence-meaning, and word-meaning. In *Studies in the way of words*, 117–143. Cambridge: Harvard University Press.
- Grice, Paul. 1989d. Meaning. In *Studies in the way of words*, 213–223. Cambridge: Harvard University Press.
- Grice, Paul. 1989e. Meaning revisited. In *Studies in the way of words*, 283–302. Cambridge: Harvard University Press.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3). 175–204.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69. 274–307.
- Gundel, Jeanette K., Michael Hegarty & Kaya Borthen. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language, and Information* 12. 281–299.
- Haviland, John B. 2003. How to point in Zinacantán. In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 139–169. Mahwah: Lawrence Erlbaum Associates.
- Heck, Richard G. 2002. Do demonstratives have senses? *Philosopher’s Imprint* 1–33.

- Heck, Richard G. 2013. Semantics and context dependence: A Strawsonian account. Forthcoming in Alexis Burgess and Brett Sherman (eds.), *New Essays on Metasemantics*.
- Heim, Irene. 1982. *On the semantics of definite and indefinite noun phrases*. Amherst: University of Massachusetts dissertation.
- Higginbotham, James. 2002. Competence with demonstratives. *Philosophical Perspectives* 16: *Language and Mind* 1–16.
- Hobbs, Jerry R. 1990. *Literature and cognition*. Stanford: Center for the Study of Language and Information.
- Isard, Stephen. 1975. Changing the context. In Edward Keenan (ed.), *Formal semantics of natural language*, vol. 287–296, Cambridge University Press.
- Jeshion, Robin. 2013a. Names not predicates. Forthcoming in *Erkenntnis*.
- Jeshion, Robin. 2013b. The semantic argument against predicativism. Forthcoming in Andrea Bianchi, ed. *New Essays on Reference*.
- Kaplan, David. 1978a. Dthat. In Peter French, Theodore Uehling & Howard Wettstein (eds.), *Contemporary perspectives in the philosophy of language*, 383–400. Minneapolis: University of Minnesota Press.
- Kaplan, David. 1978b. On the logic of demonstratives. *Journal of Philosophical Logic* 8. 81–98.
- Kaplan, David. 1989a. Afterthoughts. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 565–614. Oxford: Oxford University Press.
- Kaplan, David. 1989b. Demonstratives. In Joseph Almog, John Perry & Howard Wettstein (eds.), *Themes from Kaplan*, 481–563. Oxford: Oxford University Press.

- Kaplan, David. 1990. Words. *Proceedings of the Aristotelian Society* 64. 93–119. Supplementary Series.
- Katz, Jerrold J. 1994. Names without bearers. *Philosophical Review* 103(1). 1–39.
- Key, Mary. 1962. Gestures and responses: A preliminary study among some indian tribes of bolivia. *Studies in Linguistics* 16(3/4). 92–99.
- King, Jeffrey C. 1999. Are complex ‘that’ phrases devices of direct reference? *Noûs* 33(2). 155–182.
- King, Jeffrey C. 2001. *Complex demonstratives: A quantificational account*. Cambridge, Massachusetts: The MIT Press.
- King, Jeffrey C. 2008. Complex demonstratives, QI uses, and direct reference. *Philosophical Review* 117(1). 99–117.
- King, Jeffrey C. 2013a. Speaker intentions in context. Forthcoming in *Noûs*.
- King, Jeffrey C. 2013b. Supplementives, the coordination account, and conflicting intentions. Manuscript.
- King, Jeffrey C. & Jason Stanley. 2005. Semantics, pragmatics, and the role of semantic content. In Zoltán Szabó (ed.), *Semantics vs. pragmatics*, 111–164. Oxford: Oxford University Press.
- Kripke, Saul. 1972. *Naming and necessity*. Cambridge: Harvard University Press.
- Kripke, Saul. 1977. Speaker’s reference and semantic reference. In Peter French, Thomas Uehling & Howard Wettstein (eds.), *Midwest studies in philosophy, vol. 2*, 255–276. Minneapolis: University of Minnesota Press.
- Lascarides, Alex & Matthew Stone. 2009. A formal semantics of gesture. *Journal of Semantics* 393–449.

- Lawlor, Krista. 2005. Confused thought and modes of presentation. *Philosophical Quarterly* 55(218). 21–36.
- Lawlor, Krista. 2007. A notional worlds approach to confusion. *Mind and Language* 22(2). 150–172.
- Lepore, Ernest & Kirk Ludwig. 2000. The semantics and pragmatics of complex demonstratives. *Mind, New Series* 109. 199–240.
- Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.
- Lewis, David K. 1975. Languages and language. In Keith Gunderson (ed.), *Language, mind, and knowledge. Minnesota studies in the philosophy of science, vol. 7*, 3–35. Minneapolis: University of Minnesota Press.
- Lewis, David K. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(3). 339–359.
- Lewis, Karen. 2012. Discourse dynamics, pragmatics, and indefinites. *Philosophical Studies* 158. 313–342.
- MacKay, Alfred F. 1968. Mr. Donnellan and Humpty Dumpty on referring. *Philosophical Review* 77(2). 197–202.
- Matushansky, Ora. 2006a. Call me an ambulance. In Leah Bateman & Cherlon Ussery (eds.), *Proceedings of NELS*, vol. 35, 419–434.
- Matushansky, Ora. 2006b. Why rose is the rose. In Olivier Bonami & Patricia Hofherr (eds.), *Empirical issues in syntax and semantics*, vol. 6, 285–308.
- McGinn, Colin. 1981. The mechanism of reference. *Synthese* 49. 157–186.
- Michaelson, Eliot. 2013. Shifty characters. Forthcoming in *Philosophical Studies*.

- Millikan, Ruth G. 2000. *On clear and confused ideas*. Cambridge: Cambridge University Press.
- Montague, Richard. 1974. Pragmatics. In Richmond H. Thomason (ed.), *Formal philosophy*, 95–118. New Haven: Yale University Press.
- Mount, Allysoun. 2008. Intentions, gestures, and salience in ordinary and deferred demonstrative reference. *Mind and Language* 23(2). 145–164.
- Neale, Stephen. 1992. Paul grice and the philosophy of language. *Linguistics and Philosophy* 15. 509–559.
- Neale, Stephen. 1993. Term limits. In James Tomberlin (ed.), *Philosophical perspectives 7: Language and logic*, 89–114. Atascadero: Ridgeview.
- Neale, Stephen. 2004. This, that and the other. In Marga Reimer & Anne Bezuidenhout (eds.), *Descriptions and beyond*, 68–182. Oxford: Oxford University Press.
- Nunberg, Geoffrey. 1995. Transfers of meaning. *Journal of Semantics* 12. 109–132.
- Perry, John. 1977. Frege on demonstratives. *Philosophical Review* 86(4). 474–497.
- Poyates, Fernando. 1983. *New perspectives in nonverbal communication: Studies in cultural anthropology, social psychology, linguistics, literature, and semiotics*. Oxford: Pergamon Press.
- Predelli, Stefano. 1996. Never put off until tomorrow what you can do today. *Analysis* 56(2). 85–91.
- Predelli, Stefano. 1998a. ‘I am not here now’. *Analysis* 58(2). 107–115.
- Predelli, Stefano. 1998b. Utterance, interpretation and the logic of indexicals. *Mind and Language* 13(3). 400–414.
- Predelli, Stefano. 2002. Intentions, indexicals, and communication. *Analysis* 62(4). 310–316.

- Predelli, Stefano. 2005. *Contexts*. Oxford: Oxford University Press.
- Quine, Willard V.O. 1968. Ontological relativity. *Journal of Philosophy* 65(7). 185–212.
- Radulescu, Alexandru. 2012. *The logic of indexicals*. Los Angeles: University of California, Los Angeles dissertation.
- Reichenbach, Hans. 1947. *Elements of symbolic logic*. New York: Free Press.
- Reimer, Marga. 1991a. Demonstratives, demonstrations, and demonstrata. *Philosophical Studies* 63. 187–202.
- Reimer, Marga. 1991b. Do demonstratives have semantic significance? *Analysis* 51(4). 177–183.
- Reimer, Marga. 1992. Three views of demonstrative reference. *Synthese* 93. 373–402.
- Roberts, Craige. 2002. Demonstratives as definites. In Kees van Deemter & Roger Kibble (eds.), *Information sharing: Reference and presupposition in language generation and interpretation*, chap. 1, 89–196. Stanford: CSLI Press.
- Romdenh-Romluc, Komarine. 2006. ‘I’. *Philosophical Studies* 128. 257–283.
- Russell, Bertrand. 1905. On denoting. *Mind* 14. 479–493.
- Russell, Bertrand. 1910. Knowledge by acquaintance and knowledge by description. *Proceedings of the Aristotelian Society* 11. 108–128.
- Ryle, Gilbert. 1949. Discussion of rudolf carnap: ‘Meaning and Necessity’. *Philosophy* 24(88). 69–76.
- Saul, Jennifer. 2012. *Lying, misleading, and what is said: An exploration in philosophy of language and in ethics*. Oxford: Oxford University Press.

- Scharp, Kevin A. 2005. Scorekeeping in a defective language game. *Pragmatics and Cognition* 13(1). 205–226.
- Schiffer, Stephen. 1981. Indexicals and the theory of reference. *Synthese* 49. 43–100.
- Schiffer, Stephen R. 1972. *Meaning*. Oxford: Oxford University Press.
- Schroeter, Laura. 2007. The illusion of transparency. *Australasian Journal of Philosophy* 85(4). 597–618.
- Searle, John R. 1969. *Speech acts*. Cambridge: Cambridge University Press.
- Sherzer, Joel. 1972. Verbal and nonverbal deixis: the pointed lip gesture among the san blas cuna. *Language and Society* 2. 117–131.
- Sherzer, Joel. 1983. *Kuna ways of speaking: An ethnographic perspective*. Austin: University of Texas Press.
- Sherzer, Joel. 1993. Pointed lips, thumbs up, and cheek puffs: Some emblematic gestures in social interaction and ethnographic context. In Robin Queen & Rusty Barrett (eds.), *SALSA: Proceedings of the first annual symposium about language and society*, 196–211. Austin: University of Texas Department of Linguistics.
- Siegel, Susanna. 2002. The role of perception in demonstrative reference. *Philosopher's Imprint* 2(1).
- Soames, Scott. 2010. *Philosophy of language*. Princeton: Princeton University Press.
- Solan, Lawrence M. & Peter M. Tiersma. 2005. *Speaking of crime: The language of criminal justice*. Chicago: University of Chicago Press.
- Sorensen, Roy. 2007. Bald-faced lies! *Pacific Philosophical Quarterly* 88. 251–264.
- Speaks, Jeff. 2013. Speaker and hearer in the character of demonstratives. Manuscript.

- Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Syntax and semantics 9: Pragmatics*, 315–332. Academic Press.
- Stevens, Graham. 2009. Utterance at a distance. *Philosophical Studies* 143. 213–221.
- Stone, Matthew. 2001. Representing communicative intentions in collaborative conversational agents, Cape Cod: *AAAI Fall Symposium on Intent Inference for Collaborative Tasks*.
- Stone, Matthew. 2003. Linguistic representation and gricean inference. *IWCS-5: Proceedings of the Fifth International Workshop on Computational Semantics* 5–21.
- Stone, Matthew, Christine Doran, Bonnie Webber, Tonia Bleam & Martha Palmer. 2003. Microplanning with communicative intentions: The spud system. *Computational Intelligence* 19(4). 311–381.
- Strawson, Peter F. 1950. On referring. *Mind* 59. 320–344.
- Szabó, Zoltán G. 2006. The distinction between semantics and pragmatics. In Ernie Lepore & Barry Smith (eds.), *The Oxford handbook of philosophy of language*, 361–390. Oxford: Oxford University Press.
- Taylor, Barry. 1980. Truth-theory for indexical languages. In Mark Platts (ed.), *Reference, truth, and reality*, 182–98. London: Routledge.
- Thomason, Richmond H. 1990. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in communication*, 325–363. Cambridge: MIT Press.
- Tomasello, Michael. 2008. *Origins of human communication*. Cambridge: MIT Press.
- Vlach, Frank. 1981. Speaker meaning. *Linguistics and Philosophy* .

- Wettstein, Howard K. 1984. How to bridge the gap between meaning and reference. *Synthese* 58(1). 63–84.
- Wilkins, David P. 2003. Why pointing with the index finger is not a universal (in sociological and semiotic terms). In Sotaro Kita (ed.), *Pointing: Where language, culture, and cognition meet*, 171–215. Mahwah: Lawrence Erlbaum Associates.
- Williams, Bernard. 2002. *Truth and truthfulness*. Princeton: Princeton University Press.
- Wilson, Mark. 2006. *Wandering significance: An essay on conceptual behavior*. Oxford: Oxford University Press.
- Wolter, Lynsey K. 2006. *That's that: The semantics and pragmatics of demonstrative noun phrases*. Santa Cruz: University of California, Santa Cruz dissertation.
- Wolter, Lynsey K. 2009. Demonstratives in philosophy and linguistics. *Philosophy Compass* 4(3). 451–468.