

OXFORD



The Philosophy of David Kaplan

Joseph Almog
Paolo Leonardi

The Philosophy of David Kaplan

This page intentionally left blank

The Philosophy of David Kaplan

EDITED BY

Joseph Almog and Paolo Leonardi

OXFORD
UNIVERSITY PRESS
2009

OXFORD

UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2009 by Oxford University Press, Inc.

Published by Oxford University Press, Inc.

198 Madison Avenue, New York, New York 10016

www.oup.com

Oxford is a registered trademark of Oxford University Press.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or
transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

Library of Congress Cataloging-in-Publication Data

The philosophy of David Kaplan / edited by Joseph Almog and Paolo Leonardi.

p. cm.

ISBN 978-0-19-536788-1

1. Kaplan, David (David Benjamin), 1933– 2. Philosophy, American—20th century.

I. Almog, Joseph. II. Leonardi, Paolo.

B945.K354P45 2009

191—dc22 2008017069

9 8 7 6 5 4 3 2 1

Printed in the United States of America

on acid-free paper

ACKNOWLEDGMENTS

All but two essays are published here for the first time. Kit Fine's essay "The Role of Variables" appeared in the *Journal of Philosophy* 50, no.12 (2003): 605–631, and was reprinted in the *Philosophers' Annual* (2003). Sten Lindström's "Possible Worlds Semantics and the Liar: Reflections on a Problem Posed by Kaplan" appeared in A. Rojszszczak, J. Cachro, and G. Kurczewski, eds., *Philosophical Dimensions of Logic and Science*, Synthese Library, vol. 320 (Dordrecht: Kluwer, 2003).

This page intentionally left blank

CONTENTS

PART I: LIFE AND WORK

- 1 David Kaplan: The Man at Work, 3
Joseph Almog
- 2 David Kaplan: Formal Aspects of His Work, 11
C. Anthony Anderson
- 3 Quantifying into the Unquantifiable: The Life and Work of David Kaplan, 25
Nathan Salmon

PART II: LOGIC AND SEMANTICS

- 4 Higher-Order Senses, 45
Terence Parsons
- 5 Some Logical Properties of Natural Language Quantifiers, 60
Edward L. Keenan
- 6 The Lesson of Kaplan's Paradox about Possible World Semantics, 85
C. Anthony Anderson
- 7 Possible Worlds Semantics and the Liar: Reflections on a Problem Posed by Kaplan, 93
Sten Lindström
- 8 The Role of Variables, 109
Kit Fine

PART III: LANGUAGE AND THOUGHT

- 9 Reference, Inference, and the Semantics of Pejoratives, 137
Timothy Williamson

- 10** Frege's Hierarchy: A Puzzle, 159
Christopher Peacocke
- 11** Directing Intentions, 187
John Perry
- 12** Can Frege Pose Frege's Puzzle? 202
Stavroula Glezakovs
- 13** A New Starting Place for the Semantics of Belief Sentences, 208
Erin L. Eaker
- 14** What Is *De Re* Belief? 233
Robert Stalnaker
- 15** Five Theses on *De Re* States and Attitudes, 246
Tyler Burge
- Index, 317

Part I

LIFE AND WORK

This page intentionally left blank

CHAPTER 1

David Kaplan

The Man at Work

Joseph Almog

I will not write about the essays of this volume dedicated to the work of the man; the essays will speak for themselves. Nor will I write here of the man's own papers soon to appear in sequel volumes. They too will speak for themselves. I will speak rather of the man himself.¹ Not that he did not, does not, or will not speak for himself. For now fifty-six years in the business, the man has rarely not spoken. And through and through, he has spoken with unique brilliance, eloquence, and wit. Still, like many a brilliant performer in this variant show business, the man has his unspoken secrets.

He is often compared to his distinguished formally minded teachers from the fifties—Alonzo Church, Rudolf Carnap, and Richard Montague. He is also often said to be a mind-mate of his intellectual fellow travelers from the sixties—Keith Donnellan, Saul Kripke, and Hilary Putnam, all, as the sixties were simmering to a close, undoers of Gottlob Frege and creative innovators in the philosophy of language.

No doubt, forceful teachers shape their favorite students. And the lively late sixties did affect the period's intellectual rebels. Still, our man is one of a kind—a very *natural* kind, as natural as they get. A born performer, the man has been a quasi-professional magician in his spare time and has often injected tricks of that trade into his oral philosophical “demonstrations.”

Guided in the late fifties by Church, Carnap, and Montague, David got from the trio the eye for elegant formal codifications. From the last two, he inherited the *penchant* for using set theoretic model theory to represent whatever, in the world and beyond it, is so representable. He never quite thought, like Montague, that “philosophy is a definitional extension of set

1. In thinking about the man Kaplan, I have been helped by T. Burge, B. Herman, S. Shiffrin, T. Martin, E. Eaker, M. Zelikowsky, S. Coolidge, K. Goldring, N. Goldring, and P. Leonardi.

theory,” but the index sets, model structures, functions (on functions . . .), and so forth are always stashed discretely in his miniature James Bond toolbox. Most important, he got from the trio the tolerance for a plurality of philosophical approaches to a given subject complemented by an absolute intolerance for any obfuscation, equivocation, or rhetorical flourish. A student of the trio at UCLA in the late fifties would let a hundred philosophical flowers bloom, by all means, but would allow none to have a fractal geometry.

In the mold of his teachers but still a kind apart. David dissected, admired, and adopted as the ground philosophical framework the work of the German logician Gottlob Frege. His 1964 dissertation is a comparative study of the relative advantages of a Church-ed axiomatic Fregean Sinn-based logic versus those of a Carnap-ed model theoretic (“possible world semantic”) analysis. With David, as with his teachers, it was always Frege to the hilt. But unlike his teachers, when the time came, and it was a mere five years later, David was to isolate—by listening to “arguments by ear” from ordinary language—powerful data to undo his intellectual mentor, Frege.

Where Frege saw such vernacular data about our person-to-person variable associations with a name like “Aristotle” or about the behavior of indexicals and tenses as a distraction away from “demonstrative science,” David saw a window of opportunity for a *new* demonstrative science—the distinct science of semantics-of-ordinary-language. In this, David followed a line championed by his teacher, and eventually colleague and friend, the late Richard Montague. We treat English *as* a formal language, give it its *own* precise syntax and semantics, instead of *regimenting* it by a formal language.

His instinct to see-as-is the phenomenon in situ (before formalizing it in vitro) and to hear—*listen to*—our ordinary language enabled David to escape the dilemmas of the age he grew up in. Upcoming philosophers of language had then to choose between (1) Strawson’s and Austin’s ordinary language but without an exact logic and (2) Quine’s and Carnap’s exact logic but with no ordinary language. In David’s work there was the sense of mutual enrichment rather than dilemmatic choices. David would work for days in his 31 and 32 Intro to Symbolic Logic classes to find the most ingenious first-order logic symbolization of a trickster English sentence (argument). But he was also the one to show, in his theoretically important results, how plural determiners like ‘most’ and mischievous sentences like ‘Some critics admire only one another’ do escape the first-order lingua franca.

Just such a combination of arguments by ear and formal designs characterizes what may well be David’s most influential work, *Demonstratives*. Therein emerges his pioneering account of “direct reference” and what is essentially a two-stage theory of meaning. We first set our indexical/demonstrative sentence (“I am happy”) in a *context of use*. Only then, with its content fixed, do we evaluate it for a truth value in a possible world.

By the late sixties, it was clear to David, as it was to Keith Donnellan, Saul Kripke, and Hilary Putnam, that the Fregean mechanism of reference to objects and thinking about them is deficient. Each had his own way of shifting the Fregean paradigm—Donnellan with his *referential* definite descriptions (“the man drinking a martini”), Kripke and his modally rigid proper names (“Nixon”), and Hilary Putnam with his “extension-determines-meaning” natural kind terms (“Elm,” “Beech”). What was distinctive of David was not only his paradigm, demonstratives. To accommodate this paradigm in formal semantics, David was led to his most creative innovation. He isolated a new semantic factor—the *context of use*.

The embedding of the intuitive arguments by ear about indexicals and demonstratives in an exact model theory had this major benefit—a general formal theory of Meaning. Meanings emerge as rules operating on *contexts* to produce what-is-said. The meanings were called “characters” by David.

A formal theory of meaning is a virtue of David’s model theoretic approach but without the reductive vice(s) of previous formalizations. Before David and after him, there were attempts to use the possible worlds *modal* model theory as a linguistic semantics, a theory of Meaning (what the late David Lewis, also at UCLA at the time, called “General Semantics”). In 1947, Carnap wrote a book called *Meaning and Necessity*; in 1970, Kripke gave famous lectures that were to become later a book called *Naming and Necessity*. In both, the nonmodal ideas of meaning and naming (really: Kripke’s rigidity) are re-presented by ideas embedded in the modal framework, by functions on the spectrum of possible worlds.

This culminated in the late seventies in a popular generalization of David, by Bob Stalnaker and David Lewis, called “two-dimensional semantics.” This generalization enforces a double dose of modal information. The possible worlds are used both in the analysis of meaning and in the analysis of necessity. Both notions—meaning and necessity—are set theoretically reduced. To figure both, we compute *extensions* across spectra of possible worlds.

When David came up with his new paradigm, demonstratives, he was very much aware of such a two-dimensional modal generalization. His own students (and Montague’s and Dana Scott’s, who was up at Stanford and part of the model theory-it-all gang), Krister Segerberg, Hans Kamp, Frank Vlach, and the late Michael Bennett, were goading him to generalize and generalize and generalize. David recognized the *elegance* of the generalizations, but he also knew where to stop. The generalizations were breeding what, from the point of view of ordinary language, were mutants, what David called “monsters.” As Bas van Fraassen, who in the seventies was close by at USC and a visitor to David’s seminars, was to say in comparing David’s “mere” *weak* indexical semantics with all these *stronger* two-dimensional systems, David had a sense of the need to stop where ordinary language did not anymore underwrite the semantic project.

I too vividly recall this “know where to stop” sermon from David, as a twenty-one-year-old virgin at Oxford. When David gave his Locke lectures, I was driving around with him in his car, he doing errands and I enthusiastically describing to the master more and more extensions of his indexical semantics, making everything (proper names, kind terms, etc.) emerge as two-dimensional and indexical. David, his eyes to my dismay not at all on the road, was piercing me with his killer smile and saying: “Joseph, you are an eager young man, but you will learn the trick is sometimes *not* to do what we *can* formally do.”

Faithful to the linguistic data from ordinary language, David insisted, against his own apostles, that the new semantic ‘index’ is not a possible world at all. The new factor is what it is and not another thing—it is just a *context of use*. By so behaving like an intelligent superpower (not projecting all its power) David was led to major results—the separation of genuine direct *reference* from mere modal rigidity (constant *denotation* in all possible worlds). In tow, he brought about the segregation of logical truth from modal necessity. And, most important, David shows us how to disentangle the two interlaced Ms—Meaning and Modality—of Carnap’s and Kripke’s and the two-dimensionalists’ accounts.

Much more like his ancient masters, Frege and Russell, much less like the latter-day model theorists Carnap, Kripke, and the two-dimensionalists, David separates semantics and modal metaphysics. Semantics, the study of linguistic meaning, is not a branch of modal metaphysics. Conversely, a by-product of the separation is that metaphysics—for example, the study of essential properties and of what truths are necessary—is not reduced to a cryptosemantical investigation. For in David’s generation, in the sky-is-the-limit period of the possible worlds model theory (as a semantics and a metaphysics and . . .), a whole clan of metaphysicians lived and breathed inside the artifacts of the semantic models.

To ground the idea that the biological origin of Queen Elizabeth II was essential to her, one witnessed the marshaling of rigid designators, possible worlds, stipulations about who is referred to in what world, and stipulations about properties (symmetry!) of some clever binary (“accessibility”) relation on the space of points in the modal algebra. In contrast, for David, the worry is about what is *essential of her*, and it belongs in its natural home—the metaphysics of (Aristotelian) essentialism.

I speak of a solid body of doctrine and distinctions as if it were set in stone from the first day of David’s philosophical creative work. Not so. I know of no U-turn in modern philosophy of language as radical as that of David in between the very Fregean (and, at that, elegant) “Quantifying In” (1968) and the exuberantly anti-Fregean *Dthat* (1970), except for that of Bertrand Russell in between his *Principles of Mathematics* (1903) and his *On Denoting* (1905). The vivacity, the verve, the ingenuity of the critiques are similar. With one difference—by 1905, Russell takes it out on Frege, Meinong, and only codedly on his former 1903 self. In 1970, David takes it out on one target only—the Kaplan of two years earlier.

Magic year, 1970. "The times they are a changing" was much sung on campuses, and the philosophical times were changing too. This moment, with the presentation of *Dthat* at the Stanford conference in September 1970, was pivotal. The conference had another revolutionary piece, soon to be a classic, Montague's "Proper Treatment of Quantification in Ordinary English." It too generated a buzz, with its set of innovative ideas, that David and another UCLA friend, Barbara Partee, had a hand in, as one can see in the ticklish end of Montague's piece. Unlike Montague, David would not let the publishers have *Dthat*. But word was getting around that something fundamental was happening in the theory of reference.

Happening suddenly and all at once—Kripke's *Naming and Necessity* Princeton lectures of January and February 1970, Montague's innovations, Keith Donnellan's historical theory of reference, and David's shepherding of Keith's hiring by UCLA that year. The outside world was not standing still either—there were political storms engendered by the fight over academic freedom and David's fearless stand over a fair treatment of political activist Professor Angela Davis. His promotion was iced by the UC regents and the protests by eminent logicians and philosophers rained in. Every day was bringing with it another drama.

In the midst of all this, David had the time, coolness of mind, and intellectual confidence to see that his mentor Gottlob Frege, his teachers, Carnap and Church, and his own work of 1964–1970 were all wrong at the seams, the seams of *sinn* determining reference. It was quite the other way round—reference determined *sinn*, Russell's blasphemous *singular* propositions did make sense, indeed were downright inevitable; transworld identification of individuals was a pseudoproblem; quantifying into opacity—modal necessity and cognitive attitudes—made sense (and without the reductive tricks of "Quantifying In"). Just as the external world of human life and politics was undergoing one earthquake after another, David's philosophical certitudes were being shaken one after the other.

And yet, upheaval and all, *Dthat* sets out with crystalline clarity the needed reformation in the theory of reference. And launching the reformation is a simple blunt statement. *I was wrong*, says David.

It is this intellectual and moral force to say "I was wrong" that gave David such standing with his students, when he came to criticize their work. For David was and is first and foremost a teacher. His best research was done teaching, walking around the sculpture garden at UCLA and waving his arms excitedly, explaining to the tutored student the first paragraph of *Über Sinn und Bedeutung*.

Often, in this or that fancy international conference, David would open by saying that just recently he had come to see his previous, well-known views were in the wrong. And he'd say, he had been shown this by "my student," Mr. or Ms. X, knowing nobody in the audience had the faintest idea who the person was (yet!). This would create a stir—who on earth is X? But this was coming out directly from inside David's inner

cognitive life—even when he was in Barcelona or Tokyo, David had, as Donnellan would say, his students “in mind,” struggling with the ideas of their dissertations, replaying to a Catalan crowd a tutorial he carried out a week before in the UCLA sculpture garden with a twenty-something young graduate. This is David’s way of identifying new theoretical theses—“Oh, right, this was the point Hans Kamp made to me in Richard’s 1967 seminar”; “This was the point Tony Anderson showed me in my end-of-year garden party”; “This was what Paul Hovda convinced me of when I was driving with him to fix my washing machine”; and so on.

Which brings us to David’s drive-in philosophical practice. True to a common Los Angeles lifestyle, David practiced advanced technology driving and combined it with advanced philosophy tutorials. On the long and winding Sunset Boulevard, sometimes at a rather generous speed on a hairpin turn, his face totally beaming and facing his interlocutor in the passenger seat, David, in the mold of a Formula 1 driver, united into one with the sports Mercedes that was an extension of his body, found time to show you the way, the way in philosophy, that is.

This sense of unbound energy keeps recurring in reports from his much younger students, who feel out of shape when undertaking their philosophical jogging with this whirlwind of a man. The one Zen master and guide he ever allowed himself, his wife, Renee, is still hoping he will encounter that onset of tranquillity that comes with the first signs of middle age.

This inexhaustible reservoir of energy coupled with both seriousness and a sense of mischief recur in all philosophical sightings of David. Tony Anderson, Alonzo Church’s student, who had David as his “merely second” teacher, tells us:

Many years later after I was out of graduate school and teaching philosophy I encountered David at a meeting of the American Philosophical Association. We talked for a while in the hotel bar with Ruth Barcan Marcus and she finally left to retire to bed. It was then about midnight. Kaplan suggested that we continue the discussion in his room. I didn’t really care to do so. At a few very rare points in my life I have completely lost my appetite for philosophy. That was one of those times. The whole subject seemed gray and boring to me just then. But I reluctantly agreed to go up to his hotel room, partly out of politeness and partly because he had been my teacher and mentor. The discussion continued for almost *three hours*, with David doing almost all the talking. He was intense and animated as usual about some abstruse philosophical item. When I finally left his room, physically and intellectually exhausted, I suddenly realized that my interest in and love for philosophy had returned!

“Physically and intellectually exhausted” he leaves them all, sometimes half his age, but nonetheless elated and grateful for the workout. Says to us Erin Eaker, a recent student of David, who took on one of David’s gourmet topics—the semantics of indirect discourse reports—and was not going to have an easy time with it:

Every single time I met with David about my dissertation he posed the same challenge to my main thesis. Every *single* time. And every single time I convinced him that he was wrong and that I was right. Every *single* time. When I went to collect his signature on the final version of my dissertation we finally had a different conversation. I was elated! Dissertation completed, my probationary period with David was finally over. Three hours into that groundbreaking conversation, he said: "So, let's see... the question I guess I'm really asking is..." I left his house with the once-final, now-penultimate version of my dissertation, dripping with red ink, in my hand. David had thoughtfully put it into a manila envelope for me. I breathed into that envelope for ten minutes before I was able to drive home.

David-Breathless, a familiar condition among students and colleagues. You are breathless. He is all smiles, champagne bubbly, never taking himself seriously, jazz lightness at every step. But as with many a natural athlete, make no mistake—the apparent facility with which Michael Jordan slam-dunks rests on years and years of arduous work. As Quine points out to David in his Schilpp volume response, with David, philosophy is a *workout*.

David was the first philosopher who brought home to me—as I am sure to many others—that philosophy is hard *physical* work. The setting was David's Oxford Locke lectures. The local notables, led by the brilliant, witty, and elegantly dressed Gareth Evans (who was to meet an untimely death a few weeks later) all sat in an arc at the higher reaches of the imposing Glubenkian auditorium, far away from the stage. A mere first-year grad student, I was sitting rather closer to the stage. On it, I saw David dressed to the nines in a three-piece suit that, he kept telling the audience, was forced on him by his mother for the distinguished occasion. And then, David started peeling off the pieces, as he was entering the lecture proper. At a certain point, he was down to a light blue shirt and a tie, and I was worried that from here things could not go much farther—at which point, David loosened up his tie, rolled the shirt's sleeves, and turned to the blackboard to write a formula. And this was my moment of revelation—David was actually, *sweating*. It dawned on me that doing philosophy is a physical activity.

Twenty-seven years later, way into an age in which even dynamos like Jack Nicholson prioritize golf over their original craft, David is still going on full steam. And what steam. Just recently, we were giving a joint course on the classics of formal semantics from the sixties and seventies. I was responsible for crafting a take-home exam. It was long, technical, and jaw-breaking in its formulations. At 7:00 A.M., reading an e-mail sent at 12:15 A.M., I found one e-mail correcting, line by line, the ambiguities of my formulations, often finding three scope ambiguities in one sentence. This was followed by an e-mail at 4:15 A.M., in which student extraordinaire David Kaplan wrote, like the freshest graduate, a full exam, back to

back, asking me whether these are the kind of answers that would get one an A.

Forever absolutely serious and forever absolutely mischievous. And for another miracle, forever young; the undergraduates following him around fold at the knees. Middle age can wait.

CHAPTER 2

David Kaplan

Formal Aspects of His Work

C. Anthony Anderson

My assignment is to comment on the formal aspects of David Kaplan's work, and I will do so.¹ When I started to read and reread his articles and think about David Kaplan and his displays of logical philosophy, I realized that the man and his work have influenced me in very many more ways than those of which I had been consciously aware.

I will discuss some of David Kaplan's work where formal logic plays an explicit or strongly implicit role, but limiting myself to those of his works that I find especially interesting. However, that does not really rule out very much.

I won't proceed chronologically. Instead, I will intersperse the discussion with various intersections of David Kaplan's world-line with my own. I'm sure David doesn't remember all those encounters, and I was a bit startled myself when I recalled the details.

I first set eyes on David Kaplan sometime in the early 1960s. I was an undergraduate in mathematics at the University of Houston, in Houston, Texas, at the time but was also a member of the philosophy club there. Several of us decided to drive to Austin, about a hundred miles distant, to hear a paper at the University of Texas by some guy named "David Kaplan" involving some of the ideas of someone named "Got-lob Frege."

Apparently David was being considered for a job at the University of Texas, being a Ph.D. fresh out of UCLA. We reckoned that he talked fast, even for a Yankee. (For us, then, there was just the South and everywhere else. Kaplan wasn't from around there; that was clear.) A bunch of guys with string ties were grilling David about his argument—an ingenious construction from his dissertation showing that on certain attractive and plausible assumptions, one should, and almost must, take the semantical value of a sentence to be its truth-value. This was not the Frege-Church-Gödel slingshot but David's own. (Note the biblical connection: David

1. The style of this paper is very informal. It was originally presented orally at a Kaplanfest at UCLA.

and the slingshot. Coincidence? I don't think so.) He mopped up the place with the cowboys. I thought that the subject of his talk, the part of it I understood, was the most interesting thing I had ever heard anyone talk about. (David would say to me much later, about the ideas of Frege and Church on intensionality, that they "would be interesting even if there weren't any people.")

Here's a very broad sketch of the argument from the dissertation ("Foundations of intensional logic.") Assume that a semantical system consisting of a language together with a semantical relation is *Fregean*—this means essentially that the semantical values of complex expressions are to be functions of those of their parts—and that the semantical relation of denotation is to be extended from names and other singular terms, to sentences. Adopt further a kind of Occam's Razor principle: Do not multiply semantical entities beyond necessity. Then if the language is otherwise well-behaved and it permits such a construction as "the x such that if p , then $x = a$ and if not- p , then $x = b$," then one must assign the same semantical values to all the true sentence and the same semantical values to all the false sentences. So, if we take sentences to denote, we are well advised to take them to denote truth-values.

This sort of argument has generalizations to other sorts of expressions and provides a convincing case that the Fregean assignments of (extensional) semantical values display a beautiful and almost logically predetermined system of semantical ideas. I, for one, still think that the parallel between complex names and their denotations and sentences and their truth-values is so striking that scientific semantics should adopt a semantical notion that spans them—whether we call it denoting or "schmenotting." Kaplan's argument drives the point home.

Kaplan's dissertation itself is his most formal solo piece. It is squarely in the Frege-Church tradition but takes into account the then-novel ideas of Carnap on modal logic. There is a detailed construction of a formalized language that eschews indirect discourse in favor of a direct-discourse, Fregean treatment of modality. The thing is done in detail for a fragment of first-order predicate logic with identity and modality, but the full extension to modal predicate logic with identity is sketched in sufficient detail so that one can see that it will work. Two interpretations are considered according to which modality is taken to be a property of sentences or as a property of propositions. Expressions in oblique contexts are translated into the Fregean language by replacing the expression by its "analogue"—a name of the expression itself under the one interpretation and a name of the intension expressed under the other. The "possible worlds" are represented in each case by models of a language with one less layer of obliquity, in this case, one less iteration of a modality. Saul Kripke was to later simplify and otherwise improve models for modal logic by taking the possible worlds to just be "points," but the idea of identifying necessity with logical validity held sway before that innovation.

Montague, in "Pragmatics and intensional logic," was later to complain about Kaplan's system that it "appeared incapable being extended in such a way as to adequately accommodate iteration of belief" or, in general, iteration of empirical properties of propositions. To the already oft-noticed objection to Montague's own system that Carnapian (or Kripkean) propositions are identical if necessarily equivalent, whereas beliefs are not so, Montague blithely replied that "the consequence in question seems unavoidable if [Carnapian] propositions are taken as objects of belief" and that "its counterintuitive character can perhaps be traced to the existence of another notion of belief, of which the objects are sentences or, in some cases, complexes consisting in part of open formulas." Yeah, right! I would say that Kaplan's construction just should not be extended to such notions as belief or knowledge. There is a point, though, in bringing in belief. It follows from "It is necessary that $5 + 7 = 12$ " together with "Immanuel believes that $5 + 7 = 12$ " that there is something that is both necessary and the object of Immanuel's doxastic state. Such inferences cannot, it seems, be treated by taking necessity to be simply a sentential operator, and one must supply a common entity to serve as the common bearer of the property of necessity and as the relatum for the believing. This pretty much forces the Frege-like treatment and rules out sets of, or characteristic functions on, possible worlds as filling the bill. And David, to my knowledge, has never argued to the contrary.

There is much else in the piece that is still of great interest. The first chapter contains a beautiful and insightful treatment of the idea of a possible language (or semantical system) in terms of formation rules and corresponding semantical relations. The idea of a formation rule is taken as the most fundamental, and the rest of the details about the syntax of the language are to be coded therein. It's a simple way to get maximum generality for the languages being discussed while automatically taking account of idiosyncrasies of syntax. I have cribbed shamelessly from those ideas in my own recent work on possible languages, and I believe that they will stand the test of time.

Another matter of continuing interest is Kaplan's isolation and clarification of the idea of a *standard name* of something. The immediate ancestor is Carnap's notion of an *L-determinate designator* from *Meaning and Necessity*. As was characteristic of many of the logical positivists, Carnap blurred the modal and the epistemological elements in the idea. But when they are separated, one sees that the epistemological side of the thing is still important but in need of further clarification. On the modal side, a standard name is a name of some entity in every possible world (Note: *Naming and Necessity* was not yet in existence). It is important to progress in epistemic logic and the explication of the epistemological element in the theory of computation to analyze the idea of an epistemically privileged standard name. A computation is something that enables you to *find something out* in a special way. And the names of, for example,

numbers, that we count as putting us *en rapport* with them are taken for granted and hidden in the metatheory of recursive function theory.

But I digress and sermonize. Anyway, Kaplan here and elsewhere emphasizes the philosophical importance of these things. I'm not sure what he presently thinks about such matters. Names don't, for him now, convey special kinds of concepts of things anyway. Here's my favorite quote from Kaplan's dissertation. His youthful good sense, uncorrupted by the senseless denoting of the Direct Reference barbarians, shines through:

"[T]he meaning of an identity sentence is not determined solely by the individuals named. Compare, for example, the meanings of $\text{Id}(\text{'Hesperus'}, \text{'Phosphorus'})$ [i.e., the identity sentence formed with these two names—CAA] and $\text{Id}(\text{'Hesperus'}, \text{'Hesperus'})$. The meanings clearly differ, although $\text{rden}(\text{'Hesperus'})$ [i.e., the denotation of the name—CAA] = $\text{rden}(\text{'Phosphorus'})$." ("Foundations of intensional logic," p. 12)

My own attitude about such examples was given clear expression later by the French postmodernist and deconstructionist philosopher Dahveed Kaplang² in his trenchant critique of Western hegemony, "Kwahn-tee-fee-yawng Awn." He there says (I translate):

My own view is that Frege's explanation, by way of ambiguity of what appears to be the logically deviant behavior of terms in intermediate [i.e., intensional] contexts is so theoretically satisfying that if we have not yet discovered or satisfactorily grasped the peculiar intermediate objects in question, then we should simply continue looking.

The dissertation also contains a very perceptive and telling critique of Church's construction in the logic of sense and denotation of concepts of functions as functions on concepts.

Another interesting item in the dissertation is Kaplan's statement of and defense of the formula:

$$N\exists\alpha\phi(\alpha) \supset \exists\alpha N\phi(\alpha)$$

as being valid.³ Here "N" means "it is necessary that." This is a principle, he rightly points out, "to which no previous modal logician has given assent." Well, it may seem obviously in error to you: It can be necessary that something or other is a phi, a winner of the lottery, say, and yet there not be anything such that it is necessary that *it* win the lottery.

Fair enough, but that is not an instance of this schema. The overlining indicates that it is a proposition that is assumed necessary in the antecedent, and it is the existence of an *individual concept* of a certain sort that is asserted by the consequent. With that understanding and taking

2. I have used phonetic spelling here for the author's name and the title to indicate their pronunciation in French.

3. Kit Fine reminded me that ϕ must be modality free.

individual concepts to be Carnapian functions from possible worlds to individuals, the thing becomes quite reasonable, indeed undeniable. As a point of historical and logical interest, the principle is one of the axioms of Saul Kripke's later complete axiomatization of the theory of individual concepts. When the concepts in question are taken in a more Fregean spirit, the principle may be somewhat more doubtful, but that is another story.

I mention one more item of interest in Kaplan's dissertation: There is a definition of truth as a property of propositions. David remarks that Tarski gave definition of truth as a property of sentences for a wide class of languages. But, "We are unaware of any precise definition of truth as a property of propositions before the present work." He goes on to wonder at the controversy over the matter given the "essential isomorphism" of the two notions.

Kaplan doesn't quite literally state the required definition but only says that the truth predicate applied to a name of a proposition shall be equivalent to the following: For every sentence, the sentence materially implies that that proposition is a concept of what the sentence denotes.

The thing is clearer when one considers the richer extended language of the dissertation containing variables ranging over both truth-values and propositions. The implied definition of truth as a property of propositions, allowing even propositional constants and variables to occupy the subject position, would be:

$$\text{Tr}(\phi) =_{\text{df}} \forall p[p \supset \Delta(\phi, p)]$$

where ϕ is an expression of type t_1 , the type of names of or variables ranging over propositions, and p is a variable of type t , the type of truth-values, and the Churchian delta predicate expresses the relation that holds between a proposition expressed by a sentence in some possible language and the *denotation* of that sentence in that language. From the definition all the instances of the following schema (K) can be proved in the logic under discussion:

$$(K) \Delta(\phi, \Gamma) \equiv (\text{Tr}(\phi) \equiv \Gamma),$$

that is: the statement that a proposition is a concept of what a sentence denotes is equivalent to the statement that the proposition is true if and only if, well, what the sentence says is the case. This shows the satisfaction of a kind of criterion of adequacy, albeit one that only takes on the sheen of evidence when the idea of sentences denoting truth-values is accepted.

There is a more direct and intuitive definition that is directly suggested by Church's first paper on the logic of sense and denotation. Let T_t be a sentence that is true and, let's say, both necessary and known a priori. For example, ' $\forall p(p \supset p)$ ', which asserts that every truth-value materially implies itself. Such a sentence is plausibly a *standard* name of truth even in the epistemological sense. Then a truth-predicate applicable to variables ranging over propositions can be defined as: $\text{Tr}(\phi) =_{\text{df}} \Delta(\phi, T_t)$,

where ' ϕ ' is again a variable ranging over propositions, a.k.a. concepts of truth-values. A proposition is true if and only if it is a concept of . . . and here we supply a name of truth, a sentence, that is recognizably so. Schema (K) *will follow* from this definition in the logic under discussion. Moreover, for every sentence Γ , the language contains a *standard* name of the *proposition* expressed by that sentence, $\bar{\Gamma}$, and every instance of the schema $\Delta(\bar{\Gamma}, \Gamma)$ is a theorem: The proposition expressed by a sentence is a concept of the denotation thereof. Hence the propositional analogue of Tarski's Schema (T), much beloved by one and all, deflationist and inflationist, is a theorem: $\text{Tr}(\bar{\Gamma}) \equiv \Gamma$.

I find this all very satisfying. Once the Fregean idea is accepted that the semantical value of a sentence is its truth-value and is denoted thereby, everything falls into place. There is an intuitively satisfactory definition of truth from an epistemic point of view and one provably adequate as applied to propositions. I see this as some further confirmation of the Frege-Church approach. Notice, what the young Kaplan did not, that there is here no restriction to any particular language—definitions of truth as a property of propositions can be given, unrelativized to any language. It is thus not really isomorphic to the definition of truth as a property of sentences as applied to any one language. And it is not easily deflatable.

Such are some of the riches of Kaplan's longest publication.

Okay, back to the interesting stuff—my life. I graduated from the University of Houston with a master's degree in mathematics, but I was already in love with philosophy. As I think back, I believe that Kaplan's talk contributed mightily to my decision to go to graduate school in philosophy. I went off to the University of Illinois, at Urbana, to study many-valued logics and a bit later, in 1967, traveled from there by train to Chicago, Illinois, for the APA convention, the first I had ever attended.

Yes, there he was, in a colloquium on oblique contexts also starring Terry Parsons and Jaakko Hintikka. Kaplan's contribution was later published as "Trans World Heirlines." Kaplan's presentation was a blast of intellectual fireworks and philosophical humor. The system Kaplan there formulates (called "Essentialism" in Kaplan 1979) again touches on matters close to my heart and mind. It has a strong Fregean flavor with Carnapian toppings—yum, delicious! The APA meeting left me starry-eyed. By the by, it was my first glimpse of Richard Montague. He gave a talk at a joint session with the Association for Symbolic Logic in which I heard him say this, almost verbatim: "Many have thought that it is impossible to formulate an adequate system of modal logic because of the criticisms of Quine and the failures of such logicians as Carnap and Church. Actually, it's quite easy." Wow, I thought. Either this guy is a colossal bluff or he's what they would call in baseball terminology a "phenom"—Montague was the latter. Anyway, I rode the train back to Urbana in a state of dazed philosophical euphoria.

Shortly my mathematics teacher at Urbana, W. W. Boone, himself a significant figure in the history of logic, would suggest that I was in the

wrong place at the University of Illinois. He said that I should go to California and study under his former teacher, Alonzo Church!! Soon I was talking to David Kaplan (I got to speak to him at last), and he invited me to come out to UCLA. We loaded up the old Peugeot, and off we went, going West, seeking philosophical gold.

At the reception for graduate students at UCLA, in the Philosophy Reading Room in Dodd Hall, I overheard Kaplan talking to one of the other faculty members. (It's not difficult to overhear David's conversations, although I was way on the other side of the room.) Rudolf Carnap was professor emeritus and had recently undergone a cataract operation. He needed, said David, someone to read to him. In two bounds I was across the room and volunteered for the job. I did go read to Carnap for a while. Later, when I first read David's account of his meeting with Carnap to discuss the positivist's criterion of meaningfulness, I remembered walking up the stairs to Carnap's apartment for those meetings and the intense feeling of awe his presence inspired. Kaplan, years before, had gone to visit Carnap as a graduate student and describes his encounter in "Homage to Carnap." David, the graduate student, had proceeded to explain to one of the undeniably greatest philosophers of the twentieth century that Carnap's revised, elaborate, and formal version of the verification criterion of meaning (Carnap 1956) falls to a sophisticated version of Alonzo Church's counterexample (Church 1949) to Ayer's earlier attempt. Carnap agreed. That occasion marks the philosophical end to logical positivism as the dominant force in philosophy in the twentieth century. Kaplan's critique was eventually published as "Significance and analyticity."

At UCLA I took courses from David Kaplan and remember especially a seminar on Essentialism. Perhaps he was then working on the ideas of "Quantifying In" (Kaplan 1968). I'm not sure about the dates. I *loved* the seminar. David had, and retains, a kind of boyish exuberance for philosophy that is extremely contagious. I think I have pretty much copied David's teaching style, although not altogether consciously. The enthusiasm needs to be real, so it's not something that can be voluntarily adopted, but I do try to convey my real love for the subject to my students.

I was at a conference in Salzburg, Austria, in 1983 where David presented an early version of "A problem in possible world semantics." David's puzzle was published in the *Festschrift* for Ruth Barcan Marcus (Kaplan 1995). I've already had my say about that at the last Kaplanfest in San Marino, so I won't go into it.⁴ Many of you know that it points out a fundamental sort of limitation on possible world semantics. The idea had already inspired a good deal of activity in the philosophical logic literature when it actually came out in print (e.g., by David Lewis,

4. Published in the present volume as chapter 6.

Michael Jubien, and Martin Davies). And it really does prove that there is a certain limitation to possible world semantics for modal logic (unrelated to those already mentioned above). It's an extremely ingenious and fundamental argument.

There is still another point at which David's work had a special importance for me. Somewhere around 1980, I was foundering about at the University of Minnesota. I was there as a two-year visitor, having left my job at the University of Texas, and was competing for a job that had opened up at Minnesota. They were especially interested in an epistemologist. Okay, I can do epistemologist. I wrote up some stuff I had been thinking about for some time on the paradox of the Knower. That paradox, you may know, was first formulated by one David Kaplan and a Richard Montague in a paper called "A paradox regained" (David's title, I'll bet). It results from an ingenious and deep analysis of the "Surprise Examination Puzzle." Here's the gist of the puzzle: The instructor of a course announces that one day next week there will a *surprise* exam—you will not be able to reason your way to the day on which the exam will be given on any day before that event. First you prove that such an exam can't be given on Friday. If it isn't given by Thursday evening, then of course you would know on Thursday that it would be given on Friday. So Friday is absolutely ruled out. Okay, it will be on one of Monday through Thursday. Well, it can't be on Thursday, because we have already proved that it can't be on Friday—and if it is not given by Wednesday evening, we would be able to anticipate it at that time. That leaves Monday, Tuesday, and Wednesday. Continuing this line of reasoning, you prove that the surprise examination cannot be given at all and are most startled when it is given on, say, Wednesday. You have proved that the surprise exam cannot be given, but in defiance of logic, it has been. What's wrong?

Kaplan and Montague, in the paper under discussion, give a careful analysis of the paradox, and by exploiting a certain implicit self-reference, they turn it into a refutation of certain apparently self-evident principles of epistemic logic—that whatever is known is true, that this is itself an item of knowledge, and the like.

Later Montague, in "Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability," would further refine the argument of the paradox into a proof about the inconsistency of certain logics that take necessity to be a predicate of sentences. (The argument does not apply to the system of Kaplan's dissertation or to a later construction by Bryan Skyrms in his "An immaculate conception of modality"—essentially, in both cases, because of the meager means available for referring to sentences.) Some have seen Montague's work as a kind of "proof" that necessity must be taken to be a sentential connective—not I, but it certainly presents another challenge for those who truck with properties of sentences or propositions. But such is already the legacy of the Liar paradox for those who do not think that truth vanishes

with the morning mist if subjected to the trivialization procedure implicit, some say, in Tarski's schema T.

I lost a lot of sleep over this thing, discarded a large number of proposed solutions, and ended up recommending a Russell-like hierarchy of knowledge predicates in my paper. (Come to think of it, this was one of the possibilities already mentioned by Kaplan and Montague.) I gave the paper at the University of Minnesota as my job talk, and it was later published in the *Journal of Philosophy*. I got the job. Kaplan and Montague's work had led to the much-to-be-desired saving of my precious and proverbial bacon.

"A paradox regained" is an early paper. David Kaplan was still a graduate student when it was published! Shortly thereafter there appeared another joint publication, this time by David, Richard Montague, and Rolf Eberle. The paper was called "Hempel and Oppenheim on explanation" and was immediately followed by David's solo piece "Explanation revisited," both in the journal *Philosophy of Science*. The beginning of modern work on explanation was the paper of the title. There it is proposed that an explanation consists of a law that, together with initial conditions, logically implies the fact to be explained. Various conditions are attached to rule out immediate counterexamples. Hempel and Oppenheim used some symbolic logic in their official definitions, as was coming to be the rule with logical positivists and their sympathizers. Too bad—if you dare to actually put things precisely, you may be found out to have made a mistake. Eberle, Kaplan, and Montague produce several different ways of constructing examples that show, in effect, that almost any true general principle can be made to explain almost any singular fact—according to the proposed analysis. They use, among other things, Craig's Interpolation Lemma, then only four years proved and not yet, I imagine, even called by that name.

Powerful stuff—theorems and techniques of symbolic logic applied to philosophical problems. I loved it. I still do. In his separate paper, immediately following in the same issue of *Philosophy of Science*, David proposed a modified definition of explanation that evades all the problems found by Eberle, Kaplan, and Montague. As far as I can find out, no one ever was able to propose counterexamples of a similar sort that would defeat David's emended definition. None is mentioned, for example, in Wesley Salmon's book *Four Decades of Scientific Explanation*. There was some criticism that the emendations were not sufficiently well motivated, but in general criticisms of the Hempel-Oppenheim model of explanation took other directions thereafter. I wonder why David didn't appeal to relevant logic, since the difficulty turns on fact that in classical logic one can infer " p or (irrelevant) q " from just p . Perhaps it is because the first paper on the topic by Belnap had just appeared the year before. Still, I don't know why David didn't just invent relevant logic on the spot for the purpose.

The results are solid. The original analysis failed and was shown to have failed, and David's repair effectively blocked all similar difficulties.

Later, Richard Montague, in his "Deterministic theories," provided definitions of "deterministic theory" and of "provably deterministic theory," among other things, and proved a certain Theorem 12 involving the definitions: namely, that predicative theories that are provably deterministic are deterministic. He wrote there, "Theorem 12 no longer holds when we consider theories whose language is L_2 [a certain first-order language]. This fact, which was discovered by Mr. David Kaplan, detracts considerably from the interest of Theorem 11. In view of this situation, it would seem profitable, when considering theories formulated in L_2 , to explore other definitions of provable determinism than the one provided by Definition 18. An interesting definition of this sort, for which Theorem 12 remains valid, has been suggested by Mr. Kaplan, but will not be discussed here." Those pesky graduate students!! They show that some of our work is not very important and then have the cheek to show how to fix it. David Kaplan thus also contributed to one of Montague's early precise papers in the philosophy of science. All this is actually before David's dissertation.

In fact, there was another joint paper with Montague called "Foundations of higher-order logic" that was published in 1965, but that I suppose was done while David was still a graduate student. (David's dissertation was completed in 1964.) Reading the paper is a real workout. It bears the distinctive marks of Richard Montague's prose and personality. The paper is couched in a "rather rich metatheory." There are four, count 'em, four levels of classes—sets, proper classes, superclasses, and, though they aren't actually separately named, I guess, super-duper classes. The authors assume the existence of a "good many individuals" too. For each pure set it is postulated that there is a set of individuals of the same cardinality. And, I quote, "It would be *quite easy* to construct an axiomatic metatheory satisfying our two assumptions." But "Neither assumption... is essential; at the expense of complicating certain constructions we could work entirely with the standard set theory of Bernays-Morse" (now often called Kelly-Morse). "The method of representation is *rather obvious*, but is given in detail in..." and here a projected book by Montague, Scott, and Tarski is named. It never appeared. The paper goes on to give definitions of truth and satisfaction for higher-order languages and to solve a problem posed by Tarski about defining isomorphism for higher order structures. I was daunted. (I had heard this story about Richard Montague: When someone said about a theorem Montague had proved, "I see that the proof is correct, but I cannot see any leading principle and I have no idea whatever how anyone could or would think of such a proof." Montague thought for a moment and replied, "Ah, yes, but that's mathematical elegance, isn't it?")

One can't tell exactly who contributed what beyond the undeniably Montagovian tone of the written version, except for one place where it is said that "Kaplan has recently, in [2], accomplished a further reduction:

with each sentence of pure dyadic second-order logic can be effectively associated a logically equivalent sentence of pure monadic *third-order* logic." The reference is to a paper by one D. Kaplan called "Relations reduced to classes," but it never appeared.

Nathan Salmon has written at length about "Quantifying In." Recalcitrant indirect-reference theorist that I am, I still like many things about that paper, although I gather that David has experienced some pangs of regret. The idea of a standard name or, as I would prefer, a standard concept surfaces again in "Quantifying In." In addition to the need for an epistemic version mentioned earlier, I still like David's way of connecting the *de re* and the *de dicto* in the case of modality. I have used it in some of my ongoing work on doing modal logic in the style of Church and Frege. Here again I am doing what one of David's former students, Harry Deutsch, poignantly described as "picking up Kaplan's leavings."

"How to Russell a Frege-Church" is a later piece in which it is sketched how one might reduce Church's logic of sense and denotation, with certain slight modifications, to a Russellian intensional logic by way of the correlation of Fregean individual concepts with Russellian unique-fying properties. The thing is brilliant in conception and execution. I still have the mimeographed notes wherein Kaplan works out the theorem in detail and remember marveling at the sweep of the idea and the sheer ingenuity and intricate detail of the construction. The idea of representing individual concepts by properties that uniquely determine an individual is not so hard to hit upon (it appears, for example, in Montague's "Pragmatics and Intensional Logic," but is in any case a fairly natural maneuver). But the paper does much more than that. There is a discussion of the important contrast concerning the metaphysics of modal logic between Haeccetism and its opposite, Anti-Haeccetism. And in constructing his version of Church's logic ("El Churcho," he calls it), he drops Church's problematic construction of concepts of functions and adds the idea of the "composition" of two intensions as a primitive of the logic. Both of these steps are, in my opinion, in the direction of truth, and I have embodied them in my recent work on intensional logic, although I have strayed from the type-theoretical structure of the logic of sense and denotation. More leavings collected.

"Opacity" contains some very interesting hints as to how one might go about constructing a *general* logic of the *de re* and the *de dicto*. It is something that is important philosophically and should certainly be worked out. I used something very like Kaplan's idea of the "valuated sentence" in my paper on Russellian intensional logic. Somehow I thought that some of my ideas actually came from me.

"The Logic of Demonstratives" makes valuable a philosophical application of the idea of double-indexing, first introduced formally by Hans Kamp and Frank Vlach in their studies of temporal logic. I suppose the fundamental intuition comes from Hans Reichenbach's discussion of tense in *Elements of Symbolic Logic*. David introduces the important

distinction between character and content and applies it in a very illuminating way to the semantics of demonstratives. Alas, others, who shall here be nameless, have used double-indexing in connection with much shakier philosophical projects in the philosophy of mind.

In some abstracts in the *Journal of Symbolic Logic*, David reports on his work on S5 with quantifiable propositional variables and on Rescher's plurality quantification. I like both of them very much. I did some work on extending the syllogistic forms to include "Most A's are B's" as another form of categorical proposition when I was still working on my dissertation and was teaching at USC. I am quite pleased to be able to report that I did not get any of the ideas of that work from David Kaplan.

Well, I haven't discussed all of David Kaplan's formal work. Some of it, like his formulation of Russellian Type Theory, remains in mimeographed notes. (Yes, mimeographed. It's been a while.) I have some copies of correspondence between Kaplan and Quine on the theory of types. David let me copy them when I was taking an independent study with him on the theory of types. Quine gets knocked about a bit, but always in Kaplan's playful and friendly way. The exchange is noted, although not thus described, in the second edition of Quine's *Set Theory and Its Logic*.

I know something about Alonzo Church's estimation of David. Near the end of my career as a graduate student, I was one of the assistant editors for the reviews section of the *Journal of Symbolic Logic*, headed by Church at UCLA. Church was very careful about assigning reviewers. If the importance or value of the piece was not matched by the reviewer proposed by an assistant editor, Church would just reject the suggestion. He had final say over all such assignments. A very competitive matter among assistant editors was to try to think of a suggestion for a reviewer for Quine's *Word and Object* who wouldn't be rejected by Church. It was common knowledge among the assistant editors that David Kaplan was the only suggested reviewer for that work that Church had ever approved. David declined. I would have loved to see the outcome of such a review. Happily, he did agree to review one of Kripke's important papers on modal logic.

One final bit of Church-Kaplaniana. To grasp the full import of this quotation, you must know that Church had excruciatingly high standards of excellence. He was not a man who threw around compliments. In his monumental *Bibliography of Symbolic Logic*, he gave high ratings—in the form of double stars—to only about six works and the thing covers symbolic logic from Leibniz to the 1930s! Church gave severely accurate evaluations of some very important philosophers and logicians—never mean, but always applying a standard to which few could measure up.

He said about David, "He is the man best qualified to continue the debate with Quine about intensionality." No one, in his opinion, was better qualified to debate one of the most influential philosophers of the twentieth century. High praise indeed. He also said to me about David's

dissertation that it was “required reading.” *Principia Mathematica* and the Bible were perhaps the other items on Church’s required reading list.

Kaplan’s philosophical work often has a formal system or its semantics either in front of or in back of the discussion. There is, however, none of the flaunting of symbols that one learns early on can be used to intimidate those who are quantificationally challenged. The formalism is used for what it is good for—to crystallize ideas and to test their consequences. This is the true method of analytic philosophy. I am very grateful indeed to the powers that be for David’s existence and for the deep and stimulating works he has produced.

BIBLIOGRAPHY

- Anderson, C. Anthony. 1983. “The paradox of the Knower,” *Journal of Philosophy* 80: 338–355.
- Anderson, C. Anthony. 1989. “Russellian intensional logic.” In *Themes from Kaplan*, ed. Joseph Almog, John Perry, and Howard Wettstein, 67–103. Oxford: Oxford University Press.
- Carnap, Rudolf. 1956. “The methodological character of theoretical concepts.” In *The Foundations of Science and the Concepts of Psychology and Psycholanalysis*, ed. Herbert Feigl and Michael Scriven, 38–76. Minneapolis: University of Minnesota Press.
- Carnap, Rudolf. 1960. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: Phoenix Press.
- Church, Alonzo. 1949. “Review of A. J. Ayer, *Language, Truth, and Logic*,” *Journal of Symbolic Logic* 14: 53.
- Church, Alonzo. 1951. “A formulation of the logic of sense and denotation.” In *Structure, Method, and Meaning: Essays in Honor of Henry M. Sheffer*, ed. Paul Henle, Horace M. Kallen, and Susanne K. Langer, 3–24. New York: Liberal Arts Press.
- Eberle, Rolf, David Kaplan, and Richard Montague. 1961. “Hempel and Oppenheim on explanation,” *Philosophy of Science* 28: 418–428.
- Kaplan, David. 1961. “Explanation revisited,” *Philosophy of Science* 28: 429–436.
- Kaplan, David. 1964. “Foundations of intensional logic,” Ann Arbor, Mich.: Ann Arbor Microfilms.
- Kaplan, David. 1966. “Review of Saul Kripke, ‘Semantical analysis of modal logic’,” *Journal of Symbolic Logic* 31: 120–122.
- Kaplan, David. 1968. “Quantifying In,” *Synthese* 19: 178–214. Reprinted in L. Linsky, ed., *Reference and Modality*. New York: Oxford University Press, 1972.
- Kaplan, David. 1975a. “Significance and analyticity: A comment on some recent proposals of Carnap” and “Postscript 1975.” In *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*, ed. J. Hintikka, 87–94. Dordrecht, Holland: Reidel.
- Kaplan, David. 1975b. “How to Russell a Frege-Church,” *Journal of Philosophy* 72: 716–729.
- Kaplan, David. 1979a. “Trans World Heirlines.” In *The Possible and the Actual: Readings in the Metaphysics of Modality*, ed. M. Loux, 88–109. Ithaca: Cornell University Press.

- Kaplan, David. 1979b. "On the logic of demonstratives," *Journal of Philosophical Logic* 8: 81–98.
- Kaplan, David. 1986. "Opacity." In *The Philosophy of W. V. Quine*, ed. L. Hahn, 229–289. Chicago: Open Court.
- Kaplan, David. 1995. "A problem in possible world semantics." In *Modality, Morality, and Belief; Essays in Honor of Ruth Barcan Marcus*, ed. Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher, 41–52. Cambridge: Cambridge University Press.
- Kaplan, David, and Richard Montague. 1960. "A paradox regained," *Notre Dame Journal of Formal Logic* 1: 79–90. Reprinted in Thomason, *Formal Philosophy*, 1974.
- Kaplan, David, and Richard Montague. 1965. "Foundations of higher-order logic." In *Logic, Methodology and Philosophy of Science: Proceedings of the 1964 International Congress*, ed. Y. Bar-Hillel, 101–111. Amsterdam: North-Holland.
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Montague, Richard. 1962. "Deterministic theories." In *Decisions, Values, and Groups 2*, ed. Norman F. Washburne, 325–370. Oxford: Pergamon Press. Reprinted in Thomason, *Formal Philosophy*, 1974.
- Montague, Richard. 1963. "Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability," *Acta Philosophica Fennica* 16: 153–67. Reprinted in Thomason, *Formal Philosophy*, 1974.
- Montague, Richard. 1970. "Pragmatics and intensional logic," *Synthese* 22: 68–94. Reprinted in Thomason, *Formal Philosophy*, 1974.
- Montague, Richard. 1974. *Formal Philosophy*, ed. R. H. Thomason. New Haven, Conn.: Yale University Press.
- Quine, W. V. O. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.
- Quine, W. V. O. 1963. *Set Theory and Its Logic*. Cambridge, Mass.: Harvard University Press.
- Salmon, Wesley C. 1989. *Four Decades of Scientific Explanation*. Minneapolis: University of Minnesota Press.
- Skyrms, Brian. 1978. "An immaculate conception of modality, or, how to confuse use and mention," *Journal of Philosophy* 75: 368–87.
- Thomason, R. H., ed. 1974. *Formal Philosophy*. New Haven, Conn.: Yale University Press.

CHAPTER 3

Quantifying into the Unquantifiable

The Life and Work of David Kaplan

Nathan Salmon

I

We are gathered here to celebrate a happy milestone in the life of a special friend, David Benjamin Kaplan. Exactly one month ago, on September 17, the Earth had journeyed round the Sun exactly seventy times since the moment of his birth. Although we don't think of birthdays this way, it can be consoling to recognize that one's birthday constitutes what philosophers call a *merely Cambridge change*. Like Xanthippe's becoming a widow when the hemlock took its effect on Socrates, it is a nonintrinsic change in status in virtue of an intrinsic or "real" change in some related thing (in the case of a birthday, the Earth's periodic orbits since the time of birth). It is somewhat less comforting to acknowledge that, like Xanthippe's becoming a widow, this type of merely Cambridge change also marks an accumulation of intrinsic changes. David has a great many non-merely-Cambridge properties that are equally celebration-worthy. It is my honor to provide a brief survey.

I once heard a speaker say, "Before I begin my talk, I would like to say a few words." When I was asked to deliver a keynote address about David's life and work, I knew I could talk for hours about his work, but I wondered at first if there would be very much to say about his life. David's life is the Platonic Form of stability. He has been blissfully married to his college sweetheart for nearly fifty years, has been at UCLA for fifty-two years, and has lived in Los Angeles nearly his entire life. But there is much to say about his life—too much, in fact, for me to fit it all in today. David's work is only a part of his story. His life has been a rags-to-riches tale. From humble beginnings emerged a unique form of greatness not only as a philosopher but as a multifaceted, multitalented

The present essay was delivered on October 17, 2003, as the keynote address at DavidFest, a seventieth birthday celebration at UCLA in honor of David Kaplan. I thank my audience for their warm reactions, and Gary Mar for his assistance in devising a title.

thinker, a pundit, a comic genius, a family man, a man of principle, a man of observation and a man of action, a man with a tremendous intellect and a heart to match. David is, quite simply, a force of nature. With all his energy and enthusiasm, it is hard to believe that he is even middle-aged, let alone advanced middle-aged. As long as he is alive, the UCLA philosophy department is in no danger of an energy crisis.

I begin by talking about another of David's merely Cambridge changes. To the best of my recollection, my first encounter with David was thirty-two years ago, during winter quarter 1972, when, on the encouragement—or, more accurately, at the insistence—of Harry Deutsch, I took David's course in intermediate logic. Harry had been my instructor in introductory logic during the preceding term. That was my first term at UCLA, but already I sensed that there was something very special, almost magical, going on here, especially within the logicians' wing of the department. I had completed two years as a math major at a respectable community college—including a semester-long "baby logic" course on truth tables and Aristotelian syllogisms, and another course on logical positivism taught by a former student of Hans Reichenbach—and I was struck by the obviously greater intelligence I found here. My logic TA was smarter than anyone I had ever encountered before coming to UCLA, and Harry was brighter still. There were rumors floating around about a genius named Richard Montague, who had died tragically only months before, and whom I knew only by description—as coauthor with Donald Kalish of the rigorously demanding but marvelous textbook we had used in the introductory logic course.

At the end of the term, Harry called in the two students who had performed best on his final exam, to give us a kind of recruitment pep talk, to counsel us about where to go from there. He told us that Professor David Kaplan, who was scheduled to teach the intermediate logic course the following term, was one of the most eminent, and one of the best, philosophers alive. By taking Professor Kaplan's logic course, we would be prepared to take his more advanced courses, as well as courses to be given by the great Alonzo Church and by a thirty-year-old phenomenon and former child prodigy named 'Saul Kripke' who would be visiting in the spring. The names—Church, Kaplan, Kripke—I had never heard before. But my very logical logic instructor spoke with such obvious admiration for these brilliant logicians that for me to ignore his solemn advice would have been illogical.

I didn't realize it at the time, but it was perhaps the second best piece of advice ever given, by anyone to anyone. That academic year of 1971–1972 was educationally the richest of my life, owing largely to David's masterful tutelage. For the next six years, until I left sunlight for ivy, I took every course that David taught. I did the same with my other teachers: Tyler Burge, Alonzo Church, Keith Donnellan, Don Kalish, and Saul Kripke. More than anyone else, it was David who would become my mentor. Under his influence I came to see just how extraordinarily fascinating

philosophy is. It was only later that I came to the awareness that philosophy is *not* in fact a uniformly fascinating discipline. Some of you who have worked closely with David may not yet realize this yourselves, but some aspects of our discipline are in fact downright boring, or at best, an acquired taste. What was genuinely fascinating, I later discovered, was what happens when philosophical material is wrought by the hands of a true master.

Like everyone here, I have learned from David about many things, philosophical and otherwise, and I am continuing to learn from him. In the preface to my book *Frege's Puzzle*, I acknowledged my indebtedness to David "both for the many ideas that stem from his work and for the many intellectual benefits that accrue from being his friend." David told me at the time that that was the nicest thing anyone had ever said of him. I still feel a touch of sadness at recalling his remark. It is no exaggeration to say that to know David is to be a student. Perhaps those of us in David's debt have not said as much as we should have because when we're speaking with him, we're too busy taking mental notes. Also, when speaking with him, it can sometimes be a challenge to get a word in edgewise. I trust this celebration will rectify that situation.

II

David wrote the following autobiographical comment: "Throughout my life, I have had the uncommonly good fortune to fall under the influence of persons of great intelligence, good humor, and tolerance. Principal among these are my wonderful parents, Martha and Irv Kaplan, my inspiring teachers, Rudolf Carnap and Donald Kalish, and my remarkable wife, Renée Kaplan, the *ne plus ultra* of all three qualities."¹

Irv and Martha were first-generation Americans, neither of whom graduated from college. As a child, their son attended a variety of schools, mostly in California, including a military school. His teachers considered him disruptive because he talked a lot. But he thrived at Palm Springs Elementary, which was a modern, unstructured school. Always politically active, when he was thirteen he worked on the congressional campaign of Jerry Voorhis against Richard Nixon. David's one abiding passion during his high school and early college career at John Muir School in Pasadena was jazz, and he played trumpet in a jazz combo. He was admitted to UCLA in 1951, but only on academic probation, owing to poor grades. His skill as a jazz trumpeter was limited as compared with that of his extremely talented peers, and he eventually gave up on his hope of becoming a professional musician. His career as a UCLA student was in serious jeopardy. With this inauspicious beginning, no one could have guessed

1. "Afterthoughts," in J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), 565n.

that David would soon be on his way to becoming a UCLA institution in his own right.

It happened by chance. His admissions adviser, Veronica, persuaded the former music major, now undeclared, to take a course taught by her husband. This was perhaps the single best piece of advice ever given, by anyone to anyone—not only because of how well it served David (others have been equally well served by their advisers) but even more because of how well it served Philosophy. Veronica's husband was Don Kalish; the course was introductory formal logic. David agreed to take the course when Veronica assured him he would love it, and that one could do well in logic even if one knew nothing about mathematics, indeed even if one knew nothing about anything. By the end of the term, David had transformed Veronica's sales pitch into a theorem.

As many here know, Don Kalish was a deep thinker and a phenomenal teacher. At UCLA he created a uniquely supportive intellectual environment for his students, one perfectly suited to David's temperament and talent. Don didn't merely instruct; he nurtured. The intellectual environment that he created valued analytical power and insight above all else, even while it strongly encouraged the efforts of students who (unlike David) didn't meet the highest standards of rigor. It was surely at least partly from Don that David acquired his skill at cultivating and polishing the raw talent and occasional diamond-in-the-rough that parade through Dodd Hall.

The die was cast. His days of playing jazz trumpet behind him, David would instead become the Miles Davis of philosophy. David took other logic-like courses. He took every course that Don taught. But there were still problems. While David excelled in his logic-oriented courses, he didn't do as well in some of his other courses. If there was one thing that the budding logician and future eminent professor couldn't care less about, it was grades. Qua student, he was either outstanding or a failure, depending on the course. His grades during his first two years here were nearly all As and Fs, with a few Cs thrown in for good measure. I still marvel at the sanguine disinterest that his uneven performance betrays.

David took a course in inductive logic from the distinguished logical positivist, Hans Reichenbach. There he met a psych major, Renée Singer, who would become his life partner, soul mate, and wife. Most of the course was nontechnical. Renée took excellent notes but didn't study much. David tutored her on the technical bits in preparation for the final exam. Renée earned an A. Her tutor, the future Hans Reichenbach Professor of Scientific Philosophy, initially received a C, until the TAs persuaded Reichenbach to give David a B. The following year, David took Reichenbach's course on probability theory. Reichenbach died midterm, and a talented Berkeley graduate student was brought in, on Don Kalish's recommendation, to fill the void. It was Richard Montague. David learned important lessons from Richard. He became a friend and, eventually, a collaborator.

I studied at UCLA throughout much of the 1970s. I found the study of philosophy here to be a serious and rigorous enterprise while, at the same

time, an immensely enjoyable experience. By that time (indeed, by the time I was born), Don Kalish had already set a special tone for the study of philosophy at UCLA—or at least for the study of logic-oriented philosophy (philosophy of language, philosophy of logic, philosophy of mathematics, philosophy of science). Richard Montague had raised the bar higher still. David continues that noble tradition. Many of us here today have had the splendid fortune of being trained in the optimal intellectual environment created by Don, Richard, David, and several others. In a labor of love, David is currently bringing the latest in technological advances to “the delightful system of boxes and cancels” (as Don called it), which Don and Richard had co-invented. David will tell you that he is doing this for a host of very important pedagogical reasons, and to make logic even more fun for students. And that is, of course, quite true. It is equally obvious, though, that he is also doing it to honor the memory of Don and Richard. It is a fitting tribute.

In 1954, Rudolf Carnap replaced Hans Reichenbach as UCLA’s senior philosopher of science, raising the bar higher still. By any standard, Carnap (as he preferred to be called) was not only a great philosopher but also one of the great thinkers of the twentieth century. And by all accounts, he was a great teacher and a great guy. Carnap’s students had included such philosophical luminaries as Carl “Peter” Hempel and Willard van Orman Quine. Quine described Carnap as “a towering figure . . . the dominant figure in philosophy from the 1930’s onward, as Russell had been in the decades before,” adding that “Carnap more than anyone else was the embodiment of logical positivism, logical empiricism, the Vienna Circle.”² And like all who studied under Carnap, Quine consistently referred to Carnap as “my greatest teacher” (p. xxiii). David, still an undergraduate in 1954, became Carnap’s student, while David and Renée became close friends of the Carnaps.

David buckled down and started taking grades a bit more seriously. Under pressure to complete his undergraduate degree—including a promise from his fiancée that they would not marry otherwise—he finally passed the courses he needed to fulfill the lower-division requirements. After four and a half years, he received his B.A. in philosophy in January 1956 and married Renée a few days later. He received the equivalent of a B.A. in mathematics the following year—incidentally, while working half-time as an electronics engineer doing logical design of computers at Hughes Aircraft. Carnap felt it would be best for David if he enrolled in graduate school elsewhere, but David couldn’t fathom the thought of leaving the riches of UCLA. After serving as a lecturer at USC in the summer of 1957, David enrolled in graduate school here, where he attended Carnap’s lectures on semantics, including his yet unpublished ideas on intensional logic.

2. “Homage to Carnap,” in R. C. Buck and R. S. Cohen, eds., *PSA 1970: In Memory of Rudolf Carnap* (Dordrecht: D. Reidel, 1971), xxii–xxiii.

In 1958, as a second-year graduate student, David studied a review by Alonzo Church of A. J. Ayer's *Language, Truth, and Logic*, in which Church decisively refuted Ayer's attempt to provide definitions that would facilitate the logical positivist's task of distilling what is empirical, and hence cognitively significant, from metaphysics-laden theories. David saw that Church's result could be strengthened using a different proof strategy. At Montague's suggestion, David applied his ideas to Carnap's much more sophisticated attempt to define notions like that of *empirical significance*, and he found that Carnap's project—which Carnap had believed in for thirty years—was indeed subject to the same sort of collapse as Ayer's. When David presented his results to Carnap, Carnap was not merely gracious in accepting the legitimacy of the criticism. He was altogether pleased by the philosophical advance, even though it came at the expense of some of his central ideas. Carnap's selfless devotion to philosophy had a tremendous impact on David. I have seen the effect of this whenever I myself have offered criticisms of this or that detail in David's work. Often he tactfully rebuts my criticism; sometimes he accepts the point. Invariably, he receives my comment with characteristic humor, grace, and appreciation.

Two years after refuting Carnap's definition of empirical significance, David demonstrated that the achievement was no fluke by doing much the same sort of thing, this time with respect to Alonzo Church, who was visiting at UCLA and lecturing on his own ideas regarding intensional logic (1960–1961). Kaplan spotted a significant oversight in the axioms of LSD (Church's Logic of Sense and Denotation). One of Church's axioms, when taken together with some obvious facts, yielded the disastrous consequence that, for any object x , there is only one Fregean concept of x —for example, only one truth, only one falsehood, and (since there are necessary truths) no contingent facts whatsoever. This refutation ultimately led to Church's Revised Formulation of the Logic of Sense and Denotation—or, as David called it, the Church Reformation—which corrected the destructive axiom. It also led to David's doctoral dissertation, "Foundations of Intensional Logic" (1964)—the last dissertation Carnap directed—which provided for the first time a set-theoretic model theory for the axiomatic logic of Fregean intensions. David's model theory could be used as a foundation for a semantics for modality. (In particular, a Kripke-style possible-worlds semantics, for example, could be constructed on this basis.)

In the grand tradition of earning a Ph.D. in philosophy at UCLA, David completed the four-year program after only eight and a half years. By this time he had already been a member of the UCLA faculty for two years. As he had with his bachelor's degree, David earned his Ph.D. under pressure, this time in order to maintain his young family's lifestyle while he would be teaching at the University of Michigan during the spring of 1964, where his salary depended on his having a Ph.D. David actually completed his dissertation while he was in Michigan,

though he had already passed his final defense the previous December. In the fall of 1964, David became assistant professor at UCLA. He was promoted to associate professor with tenure the following year. He was promoted to full professor in 1970, although the promotion was delayed because some of the UC regents at the time believed David had been recommended for promotion not for the quality of his work but on the basis of his very public opposition to the firing of Angela Davis. (He was one of the plaintiffs who won a judgment that, in effect, voided the firing.)

Since then, David has been gaining accolades and international distinction. In 1979, he advanced to a special faculty status, known at UC as "above scale" (coincidentally, at the same time that he and the other members of my doctoral committee, which he chaired, approved my dissertation). Since 1994, he has been the Hans Reichenbach Professor of Scientific Philosophy. He has served extensively in the UCLA Division of the Academic Senate, the faculty governing body. He was recently named one of the "Top 20 Professors" in UCLA's history by the publication *UCLA Today*.

III

I would like to say a word about the enduring legacy that David will leave to future generations. I can put it best by quoting another great contemporary philosopher, with whom our friend has a good deal in common. Woody Allen said, "I do not wish to achieve immortality through my work. I wish to achieve it by not dying." Like Woody Allen, David will live on through his work long after most of the rest of us are forgotten. I'll make no attempt to summarize his work. Instead, by way of an introduction to David the philosopher, I shall report some offhand remarks he has made either to me directly or in my presence. I mentally recorded these offhand remarks because I was certain they revealed something ironic, puzzling, or otherwise interesting about a brilliant philosopher. (Not to worry, David. I believe virtually none of the remarks I will quote were made confidentially.)

The first remark I will quote concerns a very particular pedagogical question. Isn't it better to teach Russell's Theory of Descriptions by having students read the chapter titled "Descriptions" from Russell's spirited monograph, *Introduction to Mathematical Philosophy*,³ rather than Russell's classic "On Denoting"?⁴ After all, the book chapter provides a more orderly presentation of the central ideas, free of the arrogant, misplaced criticisms (e.g., of Frege) and other confusions (e.g., of

3. Russell, *Introduction to Mathematical Philosophy* (London: Allen and Unwin, 1919), chap. 16.

4. Russell, *Mind* (1905), 479–493.

use and mention) that clutter the classic article from *Mind*. David's answer is a resounding "NO WAY!" "On Denoting" is the richer piece, and the pedagogically more appropriate, David insisted, in part precisely *because* of the various misplaced criticisms and confusions—even including the notorious "Gray's *Elegy*" passage, which is routinely skipped over in undergraduate courses. As an undergraduate, I had studied "On Denoting" in a course with David and also in a course with Alonzo Church and in a more advanced course with Saul Kripke, and I had found the experiences extremely instructive and illuminating. Nevertheless, on hearing David's strong preference for teaching "On Denoting," I was struck by the very idea that it is sometimes actually better to teach specific material by exposing students to historically important confusions, and correcting those confusions, instead of restricting the student's exposure to the sort of elegance and precision that one finds, for example, in Church's authoritative logic text. This idea has stuck with me, and I have tried to employ it in my own teaching (with somewhat mixed results).

Another philosophy article that, like "On Denoting," presents an opportunity to instruct by clearing away some of its confusions is Quine's classic "Quantifiers and Propositional Attitudes."⁵ In fact, although they are separated by a span of fifty years, Quine's article covers a good deal of the same conceptual terrain as Russell's, only in a more confused and biased manner—specifically, the *de-re/de-dicto* distinction, as given through quantification into and quantification within propositional-attitude contexts. David loves both of these articles and has commented extensively on each. In fact, he has contributed extensive commentaries on Quine's article to two Festschriften, nearly twenty years apart: "Quantifying In" to *Words and Objections*, edited by D. Davidson and J. Hintikka;⁶ "Opacity" to *The Philosophy of W. V. Quine*, edited by E. Hahn and P. Schilpp.⁷ These commentaries represent two very distinct periods in David's philosophical development—the earlier commentary reflecting the Fregean presuppositions he had acquired from Carnap and Church, the latter reflecting the revolutionary ideas of the anti-Fregean *direct-reference theory*, of which David is a cofounder (alongside Keith Donnellan, Saul Kripke, Ruth Barcan Marcus, Hilary Putnam, and others).

In fact—and this is the second remark I will quote—I once heard David say that he has always been interested in only one philosophical problem: how to understand quantification into an oblique context. The remark

5. Quine, "Quantifiers and Propositional Attitudes," *Journal of Philosophy* 53 (1956), 177–187.

6. Kaplan, "Quantifying In," in D. Davidson and J. Hintikka, eds., *Words and Objections* (Dordrecht: D. Reidel, 1969), 206–242; originally published in *Synthese* 18 (1968): 237–273.

7. Kaplan, "Opacity," in E. Hahn and P. Schilpp, eds., *The Philosophy of W. V. Quine* (La Salle, Ill.: Open Court, 1986), 229–289. Quine in his replies called "Quantifying In" a "masterly essay," and "Opacity" a "fun to read workout" (p. 290).

immediately struck me as nonsense. I had seen David in action many times. Especially at colloquium talks by other speakers, no one is more impressive. Even on the basis of a speaker's poorly organized presentation of half-baked ideas, David courteously and elegantly takes intellectual command of the situation—clarifying the conceptual impasse that makes the project philosophically interesting in the first place, redefining the problem, reconstructing the speaker's solution, developing it, exposing weaknesses in the solution, proposing an alternative, more satisfying solution, shedding new light on issues that remain in need of further exploration. And as far as I have been able to determine, he does this regardless of the particulars: regardless of the field, the project, and the existing literature (which anyway he hasn't read). Whether it is history of philosophy, applied ethics, philosophy of action, or the philosophical significance of ballet—no matter. It is awesome to watch, not to mention a humbling experience. How could someone do this so well whose only philosophical interest is the question of how to understand quantification into oblique contexts? I'll return to this question in a moment.

On another occasion—and this is the third remark I will quote—David expressed puzzlement to me that “Quantifying In” (1968) is still the most frequently reprinted of his articles, especially when much of his subsequent work is more significant. As I mentioned, David had produced important and influential treatises that decidedly refute the philosophical presuppositions and propensities of his earlier effort. The subsequent work included David's “Opacity,” which supersedes “Quantifying In.” David's widely studied work on demonstratives and the direct-reference theory also repudiated some of the central philosophical ideas of “Quantifying In.” “Why always with ‘Quantifying In’?” David wondered. I proposed as a possible explanation that “Quantifying In” tackles a very difficult problem for a thoroughgoing Fregean philosophy of semantics, and a general Fregean point of view will always be a very tempting stance. But my explanation was shallow, and David summarily dismissed it.

Let me try again. If there is one philosophical idea that underlies a great deal of David's work, it is that of a *singular proposition*. This is what our British colleagues call an *object-involving* or *object-dependent proposition*, that is, a proposition that is about some particular thing by virtue of that thing's occurring directly in the proposition as a proper constituent, instead of being represented therein by means of something conceptual or intensional, such as a Fregean sense (*Sinn*). As David notes, this is exactly the sort of proposition that is expressed when one uses a demonstrative appropriately in a sentence—for example, while pointing to something that is visually discernible in the context. As David also notes, it is also exactly the sort of proposition that is expressed by an “open sentence” (or open formula) under an assignment of values to its free variables.

The very idea of a singular proposition is due primarily to Russell. He had the idea well before he invented his famous Theory of Descriptions, but the idea arises quite naturally in connection with that theory's

distinctions of scope. For the *primary-occurrence* reading (i.e., the *wide-scope* reading) of 'George IV wondered whether Scott is the author of *Waverley*' positions an open sentence ('Scott is x ') within the scope of an expression of propositional attitude ('George IV wondered whether'). There is a pressing question here for the Fregean: On the primary-occurrence reading, which proposition does King George allegedly wonder about? Not so for the Russellian. If the question is raised, Russell's theory provides a ready response: The variable is a logically proper name, and the open sentence therefore expresses a singular proposition about *Waverley's* author, under the relevant assignment. The original problem of quantifying in is a Fregean problem, not Russellian.⁸

The problem took on a peculiar spin in Quine's critique. He argued that quantification into a nonextensional ("opaque") context is meaningless, in fact semantically incoherent. Suppose, for example, that of the following two sentences, the first is true, and the second false, on their Russellian secondary-occurrence (narrow-scope) readings.

- (1) George IV believes that the author of *Waverley* wrote *Rob Roy*
- (2) George IV believes that the author of *The Lady of the Lake* wrote *Rob Roy*

(King George presumably knew that Sir Walter Scott had written the poem *The Lady of the Lake*. However, Scott concealed his authorship of both *Waverley* and *Rob Roy*, while letting it be known that the same author had written both novels.) Consider now the particular construction,

- (3) George IV believes of the author of *The Lady of the Lake* that he wrote *Rob Roy*,

or what comes to the same thing, Russell's analysis of the primary-occurrence reading of (2),

- (3') $(\exists x)[(y)(y \text{ wrote } \textit{The Lady of the Lake} \equiv x = y) \ \& \ \text{George IV believes that } x \text{ wrote } \textit{Rob Roy}]$.

This construction, which involves quantification into 'George IV believes that,' is supposed to be somehow less specific, hence weaker, than either (1) and (2) in its attribution to King George. It might appear that (3) and (3') are therefore true in virtue of the truth of (1), even though (2) is false. But, Quine argued, (3) and (3') cannot even be assigned a univocal truth-value. For according to classical Tarskian semantics, (3') is true if and only if the author of *The Lady of the Lake* (i.e., Scott) satisfies the right-hand conjunct,

- (4) George IV believes that x wrote *Rob Roy*.

(Analogously, (3) is true if and only if 'George IV believes that he wrote *Rob Roy*' is true when the pronoun 'he' is used to designate Scott.) And

8. This is not to say that there are no pressing questions for Russell in connection with quantifying in.

there's the rub. For Scott neither satisfies nor fails to satisfy (4) (ditto 'George IV believes that he wrote *Rob Roy*') independently of how he is described. Scott satisfies (4) under the description 'the author of *Waverley*' (substituting the latter for 'x') but not under the description 'the author of *The Lady of the Lake*,' whereas the variable 'x' itself (and likewise the pronoun) carries with it no description whatsoever. The variable's only meaning is the individual assigned to it as value. The variable is like life: it has no more meaning than whatever meaning one chooses to give it.

Quite properly, Quine was not content with this conclusion. For King George did believe *someone* to have written *Rob Roy*—by virtue of believing that whoever wrote *Waverley* also wrote *Rob Roy*—and that someone was in fact Walter Scott. Quine proposed replacing the supposedly incoherent construction (3) with a coherent substitute, which David calls 'syntactically *de re*':

(5) George IV believes_R the author of *The Lady of the Lake* to have written *Rob Roy*.

The subscript 'R,' which stands for 'relational' (in contrast to 'notional'), indicates that the transitive verb is different from that occurring in (1) through (4). The verb for relational (*de re*) belief takes an indirect object as well as a direct object ('the author of *The Lady of the Lake*' and 'to have written *Rob Roy*'), whereas the verb for notional (*de dicto*) belief takes only a direct object ('that the author of *The Lady of the Lake* wrote *Rob Roy*'). In (2), 'the author of *The Lady of the Lake*' is within the clutches of opacity; in (5) it has been liberated, available for substitution or EG.⁹

Ostensibly, "Quantifying In" is a neo-Quinean program. Its burden is to provide a philosophical analysis of relational belief in terms of notional, that is, a philosophical definition of the syntactically *de re* verb 'believes_R' in terms of the syntactically *de dicto* 'believes.' (More accurately, the ostensible project is to provide an analysis of Quine's allegedly intension-free, antiseptic replacement for 'believes_R' in terms of his equally antiseptic replacement for 'believes.') "Quantifying In" analyzes (5) (roughly) as:

(5') $(\exists \alpha)[\alpha \text{ represents}_{\mathbf{B}} \text{ the author of } \textit{The Lady of the Lake} \text{ for George IV \& George IV believes } [\alpha \text{ wrote } \textit{Rob Roy}]]$,

where, on David's analysis, a term α **represents_B** an individual x for a subject y iff α designates x and is also a *vivid name* of x for y , in a special

9. The logical form of (5) is that of a triadic predication, as follows: **Believes_R**(George IV, the author of *The Lady of the Lake*, to have written *Rob Roy*). Quine would write (5) thus: 'George IV believes $z(z \text{ wrote } \textit{Rob Roy})$ of the author of *The Lady of the Lake*.' David formulates his version using passive voice: 'The author of *The Lady of the Lake* is believed by George IV to have written *Rob Roy*.' I cannot believe that the passive construction makes any difference. It is possible David disagrees.

sense of 'name of.' Although David does not provide a full analysis of what it is for a term (really a concept) to be a "name of" an object, he does provide a working idea.¹⁰ The important feature is that it is not merely a matter of *fit* (which might be accidental), but a matter of a *real* connection—on the analogy of a photograph being a picture of an object even if it is a terrible picture that better resembles another object. David's *representation*_B is a very special kind of designation, therewith avoiding the excesses of *latitudinarianism* (or "unrestricted exportation")—the doctrine that *de dicto* (supplemented by an existential premise) entails *de re*.

From the perspective of "Opacity," the project of "Quantifying In" is wrongheaded right from the outset. For as David shows in the later commentary, Quine's argument that (3) is incoherent is itself incorrect. Indeed, with all due respect, Russell's Theory of Descriptions together with his apparatus of singular propositions already prove that (3) is perfectly coherent. There is no need to replace (3) with (5), and hence no need to analyze (5) other than by means of (3) itself. In short, "Quantifying In" is a solution without a problem. It haggles with Quine over price, while it buys his defective bill of goods.

While this diagnosis of the situation is roughly correct as far as it goes, it misses the big picture. One need not endorse Quine's replacement of (3) by (5). Especially if one questions the philosophical propriety of singular propositions, as Frege did, one may want an analysis of (3) itself in terms of belief of Fregean "thoughts" (*Gedanken*), that is, in terms of general (nonsingular) propositions. The analysis that "Quantifying In" provides for (5) may be pressed into service for this neo-Fregean purpose.

Only now the spin is somewhat different. On Frege's view, a term cannot be assigned Scott (or anything else) as its designatum directly. Instead, the term must be assigned a sense, which independently provides an object on its own hook. For the Fregean, there are infinitely many propositions, *that the such-and-such wrote Rob Roy*, each about Scott, and none is privileged. The open sentence, '*x wrote Rob Roy*,' may equally express any one of them, but only by assigning the relevant sense-value to its free variable (the semantic analogue of substituting a description for the variable). As Russell put the matter, "there is no backward road" from Scott to any specific concept of him. The issue of which proposition is designated by '*that x wrote Rob Roy*' turns on which particular Scott-determining sense-value is assigned to the variable. In the absence of any assignment of a sense-value to its free variable, the open '*that*'-clause is without meaning.¹¹

10. Strictly speaking, on the "Quantifying In" analysis, the vivid name α is ultimately not a term but an individual concept, and the quasi quotation marks are *quasi* indirect-quotation marks. This leaves David's concept of *vividness* and being a *name of* something in need of further explanation. Still, one gets the rough idea.

11. Though they are somewhat different, Quine's argument against quantification into a nonextensional context owes more to Frege than Quine explicitly recognizes.

“Quantifying In” indirectly provides a way for a Fregean to make sense of an open ‘that’-clause without resorting to singular propositions, without assigning sense-values to its free variables, and without even regarding an open ‘that’-clause as a designating expression at all. This can be accomplished by borrowing an idea from Russell, that of a *contextual definition*. “Quantifying In” insightfully lays out a way of isolating a special subclass of propositions about Scott—let us call them the *representational thoughts*—that invoke a special sort of individual concept of Scott, one that **represents_B** Scott for George IV. For the Fregean, although ‘that x wrote *Rob Roy*’ has no meaning in isolation, it can be given a contextual definition that uses the subclass of representational thoughts while simulating the assignment of sense-values to the free variable. Specifically, the whole consisting of an open ‘that’-clause, [that φ_x] with a single occurrence of ‘ x ’ as its only free variable, occurring in an atomic sentential context,

(C) $\psi(\text{that } \varphi_x)$

is defined as an abbreviation for

(C') $(\exists \alpha)[\alpha \text{ represents}_{\psi} x \ \& \ \psi([\varphi_{\alpha}])]$.¹²

In effect, singular propositions are deemed *logical constructs* (to use an older terminology) out of representational thoughts: discourse that appears on the surface to be about a singular proposition is revealed instead, on analysis, to be about *some representational thought or other*.

Putting ‘wrote *Rob Roy*’ for φ and ‘George IV believes’ for ψ in (C) and (C'), one obtains the following as a purported analysis of (4):

(4') $(\exists \alpha)[\alpha \text{ represents}_{\text{B}} x \text{ for George IV} \ \& \ \text{George IV believes}[\alpha \text{ wrote } \textit{Rob Roy}]]$.

Plugging this in for (4) in (3') yields a trivial equivalent of (5'). We thus blaze a new trail from (3) to (5'), via Russell's (3') instead of taking Quine's unnecessary detour through (5).¹³

In the original, neo-Quinean project of “Quantifying In,” objectual quantification into the nonextensional is shunned as impermissible quantification

12. In calling (C) *atomic*, I mean that it represents an n -adic predicate together with n occurrences of syntactically appropriate terms, at least one of which is the open ‘that’-clause, [that φ_x]. The definition is generalizable in the obvious way (and in alternative ways, as David indirectly shows in “Opacity”) to open ‘that’-clauses with more than one free variable-occurrence.

13. I proposed this alternative, neo-Fregean interpretation in “Is *De Re* Belief Reducible to *De Dicto*?” in A. Kazmi, ed., *Meaning and Reference, Canadian Journal of Philosophy Supplementary Volume 23*, 1997 (Calgary: University of Calgary Press, 1998), 92; reprinted in Nathan Salmon, *Content, Cognition, and Communication* (Oxford: Oxford University Press, 2007), chap. 15. There is at least the suggestion of this idea already in “Quantifying In,” in its third footnote. Cf. also Quine, “Intensions Revisited,” in P. French, T. Uehling, and H. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press, 1979), 274 n. 9; and again “Is *De Re* Belief Reducible to *De Dicto*?” 92–93n.

into the “opaque,” while Quine’s substitute relational sense is analyzed in terms of the notional sense. On the reconstruction just suggested, apparent objectual quantification into the nonextensional is embraced, as shorthand for legitimized quantification into the “oblique” (*ungerade*), and it is explained without resorting to Russell’s singular propositions in terms of representational Fregean thoughts. Looked at in this alternative, neo-Fregean way, “Quantifying In” completes Frege’s program, filling in its most problematic lacuna in a manner that is (or purports to be) sensitive to the subtle discriminations among our ordinary *de re* attributions.¹⁴

On the neo-Fregean project, one resists Quine’s defective bill of goods, and one still gets the discounted price . . . on a different bill of goods. The new bill of goods is the rejection of singular propositions—or the demotion of singular propositions to the status of mere *logical construct*. But David has since come to like singular propositions. In fact, he loves them. So do I; I’d be nowhere without them. And neither would David. Does this mean that there is nothing of value to salvage from the project of “Quantifying In” for the later Kaplan (i.e., for the author of “Opacity”), or for the present me?

It does not. Let there be quantification into nonextensional contexts. Let there be singular propositions. Let George IV believe of the author of *The Lady of the Lake*, *de re*, that he wrote *Rob Roy*, while doubting or wondering *de dicto* whether the author of *The Lady of the Lake* wrote *Rob Roy*. Let King George do this by believing the singular proposition about the author of *The Lady of the Lake* that he wrote *Rob Roy* while doubting the general proposition. There is still a problem. For King George believes the singular proposition that Scott wrote *Rob Roy* precisely by believing a (more or less) general proposition, that the author of *Waverley* wrote *Rob Roy*. So do I, and so do you—just as Quine’s Ralph believes the singular proposition about Orcutt that he is a spy by believing that the man in the brown hat is a spy, and Kripke’s Pierre believes the

14. The original analysis purports to uncover an existential quantifier and the accompanying left-hand conjunct ‘ α *represents_B* x ’ that were allegedly concealed in (5). The analysis locates the hidden restricted quantifier ultimately in the verb ‘believes_R’ for *de re* (“relational”) belief. The neo-Fregean contextual-definition reconstruction purports to uncover the same restricted quantifier allegedly concealed in (3) and (3’), locating the hidden quantifier instead ultimately in open ‘that’-clauses. Either way, as with Russell’s contextual definitions for “denoting phrases” (determiner phrases), the restricted quantifier may compete for dominant position with other operators in more complex constructions (e.g., with the negation in ‘George IV does not believe Scott to have written *Rob Roy*,’ or with that in ‘George IV does not believe of Scott that he wrote *Rob Roy*’), with resulting scope ambiguities. Cf. section XI of “Quantifying In.”

Influenced by section XI of “Quantifying In,” I argued in *Frege’s Puzzle* (Atascadero, Calif.: Ridgeview, 1986, 1991), chap. 8, that all belief attributions—*de dicto* and *de re* alike—involve a hidden restricted existential quantifier, though a rather different one from that posited in “Quantifying In.” I locate the hidden restricted quantifier ultimately in the verb ‘believes,’ hence in the very notion of *de dicto* belief. Cf. section IV of “Is *De Re* Belief Reducible to *De Dicto*?”

singular proposition about London that it is pretty by believing that the European city called 'Londres' is pretty. But neither King George nor we believe any singular proposition by virtue of believing that the shortest spy is a spy. In particular, we fail to believe the singular proposition about the shortest spy that he or she is a spy. We cannot even apprehend the proposition.¹⁵ Nor do we believe any singular proposition by virtue of believing that the first child to be born in the twenty-second century will be born in the twenty-second century. Nor do we even apprehend any such proposition. And naming that future person 'Newman I' gets us no closer to doing so.¹⁶ There is what Derridean literary theorists would call a *difference-in-between* the two sorts of cases: Walter Scott and Orcutt, on the one hand, and the shortest spy and Newman I, on the other. What is the difference between them? Why can't we cognitively access singular propositions about the shortest spy or Newman I in the same way that we access singular propositions about the author of *Waverley*?

One possible answer is that 'the author of *Waverley*' and 'the man in the brown hat' are *representing_B* terms, whereas 'the shortest spy' and 'Newman I' are not. Question answered, problem solved. If so, "Quantifying In" shows its far-reaching vision by coming to our rescue even in the face of our acceptance of singular propositions. We have a deconstruction of "Quantifying In" and a reconstruction of it, this time neither as a neo-Quinean project nor as a neo-Fregean one, but as neo-Russellian.

Except that none of these terms can reasonably be called "vivid." ("The author of *Waverley* vivid? 'The man in the brown hat'? Give me a break.)¹⁷ David's notion of *representation_B* is, at bottom, a Fregean surrogate, or reasonable facsimile, for Russell's rather austere notion of *knowledge by direct acquaintance*. What is needed is further distance from Russell, by means of an even weaker notion of *representation*. But representation must not be weakened too much, else we stumble into the pit of latitudinarianism. I say we dump vividness and be done with it. I submit that the difference between the two sorts of cases lies in the fact that 'the author of *Waverley*' (or rather its content) is a name *of* Scott (in David's sense of 'name *of*') and 'the man in the brown hat' is likewise a name *of* Orcutt, whereas neither 'the shortest spy' nor 'Newman I' is a name *of* its respective designatum. And there we have our solution.¹⁸

15. Contrary to popular opinion, it does not follow that we cannot even assert the proposition. This is something we easily can do, even if only by introducing a name.

16. Though doing so does enable us to assert such propositions. See the previous note.

17. For this reason, the author of "Quantifying In" presumably answered Quine's puzzle question 'Is Orcutt such that Ralph believes that he is a spy?' in the negative. This despite the fact that Quine sets out his example with the explicit stipulation that Orcutt is such that Ralph believes he is a spy.

18. Cf. Keith Donnellan, "The Contingent *A Priori* and Rigid Designators," in P. French, T. Uehling, and H. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press, 1977), 45–60, sec. VI, at p. 58; and

In the course of my investigative research in preparation for this essay, I came across an internal departmental memo from the time that “Quantifying In” first appeared, in which David described the piece in these words: “‘Quantifying In’ contains almost everything (of a non-technical nature) I have thought of concerning oblique contexts over a period of six or eight years. It is my most daring piece as measured by the proportion of half-baked ideas . . .” He then adds a remarkable disclaimer, which I had always thought was invented by another of my early influences: “. . . and I would not have released it had it not been for extreme pressure from the editors” (from a December 1969 memo to the departmental chair concerning his research; David apparently failed to secure the copyright to the disclaimer).

One answer to David’s later question, “Why always with ‘Quantifying In’?” is that it has the same sort of depth and complexity as “On Denoting.” (Indeed, it has significantly more depth and complexity than “Quantifiers and Propositional Attitudes.”) Each reading brings with it new insights and conceptual gems. And even its historically important mistakes, confusions, and “half-baked ideas” have a deeper, visionary wisdom of their own. I hope David will someday find a way to come to terms with the fact that it is a great piece of philosophy.

IV

I return now to the question of why the analytical master would say that he was always interested in only one issue. There is, I think, a two-part answer. First and more obvious, David is interested in other philosophical issues besides how to understand quantification into nonextensional contexts. It is just that he is not *as* interested in those other issues.

Even this might seem a bit of an exaggeration. A quick overview of David’s most significant publications reveals a very much broader philosophical interest, as the speakers at this celebration will attest. One finds

my “The Good, the Bad, and the Ugly,” in A. Bezeudenhaut and M. Reimer, eds., *Descriptions and Beyond* (Oxford: Oxford University Press, 2004), 230–260; ; reprinted in Nathan Salmon, *Content, Cognition, and Communication* (Oxford: Oxford University Press, 2007), chap. 18.

I regret to say that, the last I’ve heard, David rejects this position, on at least two grounds. First, he evidently believes that naming Newman 1 is sufficient to tear down the cognitive wall that separates us from him or her. Cf. “Dthat,” in P. French, T. Uehling, and H. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press, 1977), 397. Second, David evidently believes that (5) says something stronger than (3). Specifically, he evidently believes that (5) requires King George to be *en rapport* with Scott after all—what I call a *de re* connection of King George to Scott—whereas (3) does not. Cf. “Afterthoughts,” in J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), 605–606 n. 95. I disagree on both counts.

an interest in demonstratives and other context-sensitive expressions, in Russell's theory of definite descriptions, in Keith Donnellan's and Kripke's accounts of ordinary proper names, in modal logic and possible-worlds semantics, in general intensional logic, even in what might be termed *philosophy of syntax*, and much more—a grab bag of philosophical topics.

More to the point, with regard to other philosophical issues, there is a sense in which the specifics really *don't* matter. There is a conceptual impasse, and there is a way to think fruitfully about the problem, maybe even to solve it. That is an end in itself. David is a born problem solver, and he is a born teacher. One might even say that he is a *compulsive* problem solver and teacher. In him, the urge to combine the two whenever possible is irresistible. And no sooner does he solve a problem than he finds an even more difficult problem, maybe even a problem lurking in the solution to the first. Instead of "If it ain't broke, don't fix it," David's credo is "If it's broke, let's fix it together; and if it ain't broke, something else is." And his motto should be, "Come on, it'll be fun!" It always is.

Finally, I would like to say a word on behalf of those of us who have studied under David. Leonardo da Vinci said, "The disciple who does not surpass his master fails him." Now there was a guy who really knew how to lay a major guilt trip on his students. Reflecting on Leonardo's words, it can be consoling to remember that, as a young man, Carnap took courses in logic and some of its philosophical applications (including logicism) from none other than Gottlob Frege. Consequently, many of us here today are not only students of David and grand-students of Carnap but also, in some sense, great-grand-students of Frege.¹⁹ We can all appreciate our intellectual pedigree. If one is going to fail to surpass one's predecessors, they might as well be unsurpassable. Also, if one is going to fail at something, one might as well do it in a very big way. I am confident that I speak for all who studied under David when I say that we feel no guilt whatsoever at having failed him by Leonardo's standards. On the contrary, it has been our privilege.

Would all of you with whom David has shared part of his life and part of himself please join me in thanking him. May he continue to be happy for many returns of the Earth to its current position relative to the Sun. Or to put it in terms of merely Cambridge change, many happy returns.

19. As if that weren't enough, Hilary Putnam has proposed that an *analytic truth* be defined as any sentence that is "deducible from the sentences in a finite list at the top of which someone who bears the ancestral of the graduate-student relation to Carnap has printed the words 'Meaning Postulate.'" "The Meaning of 'Meaning,'" in K. Gunderson, ed., *Language, Mind, and Knowledge* (Minneapolis: University of Minnesota Press, 1975), 174. Being David's student evidently carries with it certain rights and privileges—not to mention responsibilities.

This page intentionally left blank

Part II

LOGIC AND SEMANTICS

This page intentionally left blank

CHAPTER 4

Higher-Order Senses

Terence Parsons

1. INTRODUCTION

My topic today is about old-fashioned intensional logic, a topic that David Kaplan worked on a very long time ago, and one in which he is still interested. I am interested in how to develop theories of truth and meaning for natural language, focusing on the problems created by non-extensional contexts, such as that-clauses. I've worked on this for years, and I can't seem to stop.¹

A Quasi-Fregean-Carnapian-Early-Kaplanian Semantics

I assume without arguing for it here that it is a worthwhile goal of semantic theorizing to seek a theory which entails:

'S' is true iff S^*

and also entails:

the meaning of 'S' = that S^*

where the ' S^* ' is a sentence of the metalanguage that makes these equivalences obviously correct.

I have always been tantalized by whether this goal is achievable. It is clearly not achievable in general because there are many sentences of English for which it would be wrong to achieve the goal, as David Kaplan has pointed out. Examples are:

I am here now
That damn Kaplan was promoted
Oops
Clean up your room
Where is Maria?

1. I would like to thank the Padua University department for inviting me to give a series of talks on Fregean Theories of Natural Language where many of these ideas were developed.

But the question remains whether the goal might be accomplished when limited to descriptive declarative sentences without contextually sensitive elements. I used to think that even *this* is impossible, but now I think that it can be done. That is what this essay is about.

Usually, in giving a semantics for a language (the “object language”), you must use a “metalanguage” that is as rich in structure as the object language itself. I accept this; in particular, I assume that if the object language contains indirect contexts, then so must the metalanguage. Perhaps there is a way around this, but I don’t see what it is. In any event, I will presume that the metalanguage itself contains indirect contexts. I follow the spirit of Richard Montague² in using the caret ‘^’ to indicate an indirect context. An indicative sentence preceded by ‘^’ can often be pronounced ‘that S’:

^[Socrates is wise] = that Socrates is wise.

And for any grammatical category of speech, we can use a phrase to indicate its meaning; when we do this, we typically signal it with special intonation, which I represent with italics, and which I will represent in canonical notation with the meaning-caret:

‘Socrates’ means *Socrates*. ‘Socrates’ means ^[Socrates].
‘clever’ means *clever*. ‘clever’ means ^[clever].

‘Socrates is wise’ means *Socrates is wise* or *that Socrates is wise* or ^[Socrates is wise]. These examples sound trivial, which should give us confidence in their correctness. Other equally good examples do not sound trivial:

‘neve’ means *snow*. ‘neve’ means ^[snow].
‘La neve à bianca’ means *snow is white*
‘La neve à bianca’ means ^[snow is white].

I make one assumption about the “logic” of the caret: that putting a complex expression in an indirect context is the same as doing so to its parts. This is, in effect, an assumption of compositionality:

COMP: ^[F(a)] = ^F(^a)
 ^[F(a,b)] = ^F(^a, ^b)

(I use round parentheses for function-argument notation, to distinguish them from brackets, which are used merely for punctuation.) This is a schema that I take to be part of the background logic of the metalanguage.³

2. “The Proper Treatment of Quantification in Ordinary English,” in *Approaches to Natural Language*, ed. J. Hintikka, J. Moravcsik, and P. Suppes, 221–242 (Dordrecht: Reidel, 1973).

3. It’s an axiom schema, that I assume to make the theory work.

For the object language I begin by considering an artificially simple piece of English, consisting entirely of the following five items:

Vocabulary:

Names: 'Socrates', 'Mary', 'Agatha'

Predicates: 'is wise', 'believes'

The complex expressions are:

A name or names plus a predicate is a sentence.

Preceding any word or phrase by a caret (and using brackets to clarify the scope of the caret if necessary) yields a name.

For brevity, I usually abbreviate the English words and use predicate calculus word order. So examples of sentences in this notation are:

Ws *Socrates is wise*

$mB^{\wedge}[Ws]$ *Mary believes that Socrates is wise*

$aB^{\wedge}[mB^{\wedge}[Ws]]$ *Agatha believes that Mary believes that Soc is wise*

(These are the only sentences that I will be discussing for much of this essay.)

The syntax of the metalanguage is the same as the syntax of the object language, except that I will assume, with Frege, that the metalanguage has an explicit function-argument structure, and this will be indicated using parentheses. So in the object language we write

Ws for *Socrates is wise*

and this syntax leaves it open what kind of semantic structure it has. In the metalanguage the sentence will be

$W(s)$

indicating that 'is wise' stands for a function whose argument is what 'Socrates' stands for. The semantic theory will attribute to the object language a function-argument structure too, even though this is not obvious from its syntax.

In order to avoid cluttering things up with quotation marks, I use bolded versions of symbols for names of the symbols themselves. So '**F**' is a name of the symbol 'F', and '**Fa**' is a name of the string of symbols 'F', 'a'. I use Greek letters to range over strings of symbols.

Before stating the theory, it may help to have an informal illustration of how it will work. Here is how the semantical rules will apply in a sample case, using the caret notation. The following shows how the theory will establish that:

Mary believes $^{\wedge}[\text{Socrates is wise}]$ is true
if and only if believes(Mary, $^{\wedge}[\text{wise}(\text{Socrates})]$).

1. **Mary believes** $\wedge$ [**Socrates is wise**] is true if and only if **believes** is true of the references of **Mary** and of $\wedge$ [**Socrates is wise**]. (*This will be an axiom about predication.*)
2. **believes** is true of two things if and only if the first believes the second. (*This will be an axiom about believes.*)
3. The reference of **Mary** = Mary. (*An axiom about Mary.*)
4. So **Mary believes** $\wedge$ [**Socrates is wise**] is true if and only if believes(Mary, the reference of $\wedge$ [**Socrates is wise**]). (*From 1, 2, 3.*)
5. The reference of $\wedge$ [**Socrates is wise**] = the meaning of **Socrates is wise**. (*Recall Frege's observation that in indirect quotation we speak about the meaning of the embedded clause. This principle will be an axiom.*)
6. So **Mary believes** $\wedge$ [**Socrates is wise**] is true if and only if believes (Mary, the meaning of **Socrates is wise**). (*From 4 and 5.*)
7. The meaning of **Socrates is wise** = the meaning of **wise** (which is a function) applied to the meaning of **Socrates**. (*An axiom about the meaning of a predication.*)
8. The meaning of **Socrates** = $\wedge$ Socrates, and the meaning of **is wise** = $\wedge$ is wise. (*Axioms about 'Socrates' and 'is wise'.*)
9. So the meaning of **Socrates is wise** = $\wedge$ is wise($\wedge$ Socrates). (*From 7&8.*)
10. But $\wedge$ is wise($\wedge$ Socrates) = $\wedge$ [is wise(Socrates)]. (*The principle COMP introduced above; a logical principle of the metalanguage.*)
11. So the meaning of **Socrates is wise** = $\wedge$ [is wise(Socrates)]. (*From 9 and 10.*)
12. So **Mary believes** $\wedge$ [**Socrates is wise**] is true if and only if believes(Mary, $\wedge$ [is wise(Socrates)]). (*From 6 and 11.*)

Almost all of the important principles of the theory are illustrated in this little derivation.

Now I will state the theory explicitly. The background logic of the metalanguage, just illustrated, is the predicate calculus with identity, with logical laws applied to meaning names but not to expressions within them, supplemented by the principle COMP given earlier. Curly and square brackets are used for punctuation. Parentheses stand for functional application. All of the proofs are written in the metalanguage. The axioms are as follows:

ATOMIC CONDITIONS: These state references and meanings of the atomic terms:

NAMES:

ref{**Mary**} = Mary mng{**Mary**} = $\wedge$ [Mary]
 ref{**Socrates**} = Socrates mng{**Socrates**} = $\wedge$ [Socrates]

VERBS/PREDICATES:

Trueof{**Wise**, x} $\leftrightarrow$ Wise(x) mng{**Wise**} = $\wedge$ [Wise]
 Trueof{**Believe**, x,y} $\leftrightarrow$ Believe(x,y) mng{**Believe**} = $\wedge$ [Believe]

PREDICATE-ARGUMENT: These principles say how what is referred to and expressed by predicate-argument structures depend on their parts:

PRED-ARG-TRUTH: A sentence consisting of a predicate followed or flanked by its argument(s) is true iff the predicate is true of the reference(s) of its argument(s):

$$\text{True}\{fa\} \leftrightarrow \text{Trueof}\{f, \text{ref}\{a\}\}$$

$$\text{True}\{afb\} \leftrightarrow \text{Trueof}\{f, \text{ref}\{a\}, \text{ref}\{b\}\}$$

PRED-ARG-MEANING: The meaning of a sentence which consists of a predicate flanked by its argument(s) is the result of applying the meaning of the predicate to the meanings of its argument(s).

$$\text{mng}\{fa\} = \text{mng}\{f\}(\text{mng}\{a\})$$

$$\text{mng}\{afb\} = \text{mng}\{f\}(\text{mng}\{a\}, \text{mng}\{b\})$$

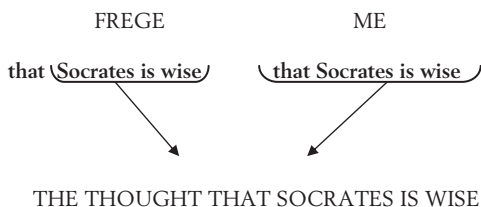
We will also need the principle of reference functionality:

$$\text{REF-FUNC} \quad \text{ref}\{\alpha\beta\} = \text{ref}\{\alpha\}(\text{ref}\{\beta\})$$

M-IDEN: Finally, we have the assumption that a meaning name, such as a that-clause, refers to the meaning that its contents express:

$$\text{M-IDEN:} \quad \text{ref}\{^{\wedge}[a]\} = \text{mng}\{a\}.$$

This is not Frege's assumption that an expression in an indirect context refers to its customary meaning; rather, it assumes that the whole indirect context refers to the meaning of its contents. In 'that Socrates is wise' it is not 'Socrates is wise' that refers to the proposition that Socrates is wise, it is the whole that-clause:



In this respect the theory is Carnapian as opposed to Fregean. The theory is thus "semantically innocent"—no word or phrase ever changes meaning or reference from one context to another. This primitive theory satisfies the material adequacy conditions for both truth and meaning for simple sentences, and it can be supplemented so as to satisfy those conditions for all sentences in our limited fragment. Here are some illustrations:

The Tarski adequacy condition for truth as applied to the sentence **Socrates is wise** can be stated as:

$$\text{True}\{\text{Wise Socrates}\} \leftrightarrow \text{Wise}(\text{Socrates}).$$

This is established from the theory in three steps as follows:

$$\begin{array}{ll} \text{True}\{\text{Wise Socrates}\} \leftrightarrow \text{Trueof}\{\text{Wise}, \text{ref}\{\text{Socrates}\}\} & \text{PRED-ARG} \\ \leftrightarrow \text{Wise}(\text{ref}\{\text{Socrates}\}) & \text{ATOMIC} \\ \leftrightarrow \text{Wise}(\text{Socrates}) & \text{ATOMIC} \end{array}$$

The Tarski criterion of adequacy for *meaning* for the sentence **Socrates is wise** requires that we be able to prove the formal analogue of:

Socrates is wise means that Socrates is wise.

This is established as follows:

$$\begin{aligned} \text{mng}\{\text{Wise Socrates}\} &= \text{mng}\{\text{Wise}\}(\text{mng}\{\text{Socrates}\}) && \text{PRED-ARG} \\ &= \wedge\text{Wise}(\wedge\text{Socrates}) && \text{ATOMIC} \\ &= \wedge[\text{Wise}(\text{Socrates})] && \text{COMP} \end{aligned}$$

More complicated cases. It is not difficult to give a theory that meets the criteria of adequacy for cases of the simple sort we have just examined. The problem has always been to include nonextensional contexts, such as sentences with that-clauses. So here is an example of the material adequacy condition for truth for 'Mary believes that Socrates is wise':

$$\text{True}\{\text{Mary believe } \wedge[\text{wise Socrates}]\} \leftrightarrow \text{Believe}(\text{Mary}, \wedge[\text{wise}(\text{Socrates})]).$$

This is proved as follows (this parallels the sample proof given earlier, but omitting extra steps that spell out substitutivity of identity):

$$\begin{aligned} &\text{True}\{\text{Mary believe } \wedge[\text{wise Socrates}]\} \\ \text{PRED-ARG-T} &\quad \leftrightarrow \text{Trueof}\{\text{believe}, \text{ref}\{\text{Mary}\}, \text{ref}\{\wedge[\text{wise Socrates}]\}\} \\ \text{ATOMIC} &\quad \leftrightarrow \text{Believe}(\text{ref}\{\text{Mary}\}, \text{ref}\{\wedge[\text{wise Socrates}]\}) \\ \text{ATOMIC} &\quad \leftrightarrow \text{Believe}(\text{Mary}, \text{ref}\{\wedge[\text{wise Socrates}]\}) \\ \text{M-IDEN} &\quad \leftrightarrow \text{Believe}(\text{Mary}, \text{mng}\{\text{wise Socrates}\}) \\ \text{PRED-ARG-M} &\quad \leftrightarrow \text{Believe}(\text{Mary}, \text{mng}\{\text{wise}\}(\text{mng}\{\text{Socrates}\})) \\ \text{ATOMIC} &\quad \leftrightarrow \text{Believe}(\text{Mary}, \wedge\text{wise}(\wedge\text{Socrates})) \\ \text{COMP} &\quad \leftrightarrow \text{Believe}(\text{Mary}, \wedge[\text{wise}(\text{Socrates})]) \end{aligned}$$

It turns out that there is only one issue remaining: to address the meanings of sentences containing that-clauses. To address it we need to say something about the *meanings* of caret-clauses. The theory as it stands does not do this. Beginning with the simplest cases we have the names:

Socrates
 $\wedge$ Socrates
 $\wedge\wedge$ Socrates
 $\wedge\wedge\wedge$ Socrates
 etc.

We must assign each of these a meaning, by adding new clauses, called ATOMIC + :

$$\begin{aligned} \text{mng}\{\text{Socrates}\} &= \wedge\text{Socrates} \\ \text{mng}\{\wedge\text{Socrates}\} &= \wedge\wedge\text{Socrates} \\ \text{mng}\{\wedge\wedge\text{Socrates}\} &= \wedge\wedge\wedge\text{Socrates} \\ \text{mng}\{\wedge\wedge\wedge\text{Socrates}\} &= \wedge\wedge\wedge\wedge\text{Socrates} \\ &\text{etc.} \end{aligned}$$

These are additional axioms—an infinite number of them for each word. (I don't much like this infinite piecemeal approach, and we will consider soon how to do better, but first we need to complete our discussion of meaning.)

We also need to generalize COMP in the metalanguage so as to apply to iterated carets:

$$\text{COMP } \wedge\wedge[\alpha(\beta)] = \wedge\wedge\alpha(\wedge\wedge\beta), \dots$$

and we add related axioms to our semantic theory (which we also call COMP), namely:

$$\begin{aligned} \text{ref}\{\wedge\wedge\wedge\wedge[\alpha\beta]\} &= \text{ref}\{\wedge\wedge\wedge\wedge[\wedge\wedge\alpha \wedge\wedge\beta]\}, \dots \\ \text{mng}\{\wedge\wedge\wedge\wedge[\alpha\beta]\} &= \text{mng}\{\wedge\wedge\wedge\wedge[\wedge\wedge\alpha \wedge\wedge\beta]\}, \dots \end{aligned}$$

These assumptions allow us to satisfy the criterion of adequacy for meaning for 'Mary believes that Socrates is wise':

$$\text{mng}\{\text{Mary Believe } \wedge\wedge[\text{Wise Socrates}]\}$$

$$\begin{aligned} \text{PRED-ARG-M} &= \text{mng}\{\text{Believe}\}(\text{mng}\{\text{Mary}\}, \text{mng}\{\wedge\wedge[\text{Wise Socrates}]\}) \\ \text{ATOMIC} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \text{mng}\{\wedge\wedge[\text{Wise Socrates}]\}) \\ \text{COMP} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \text{mng}\{\wedge\wedge\wedge\wedge[\text{Wise } \wedge\wedge\text{Socrates}]\}) \\ \text{M-IDEN} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \text{ref}\{\wedge\wedge\wedge\wedge[\wedge\wedge\text{Wise } \wedge\wedge\text{Socrates}]\}) \\ \text{COMP} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \text{ref}\{\wedge\wedge\wedge\wedge[\wedge\wedge\text{Wise } \wedge\wedge\text{Socrates}]\}) \\ \text{REF-FUNC} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \text{ref}\{\wedge\wedge\wedge\wedge[\wedge\wedge\text{Wise}]\}(\text{ref}\{\wedge\wedge\wedge\wedge[\text{Socrates}]\})) \\ \text{ATOMIC} + &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \wedge\wedge\wedge\wedge\text{Wise}(\wedge\wedge\wedge\wedge\text{Socrates})) \\ \text{COMP} &= \wedge\wedge\text{Believe}(\wedge\wedge\text{Mary}, \wedge\wedge\wedge\wedge[\text{Wise}(\text{Socrates})]) \\ \text{COMP} &= \wedge\wedge[\text{Believe}(\text{Mary}, \wedge\wedge[\text{Wise}(\text{Socrates})])]. \end{aligned}$$

It now turns out that sentences with multiple embeddings of that-clauses within that-clauses work on the model of the one just given; no additional assumptions are needed. So the pattern just given expands easily to handle both truth and meaning for, for example, 'Agatha thinks that Mary believes that Socrates is wise'.

To conclude section 1, I will comment briefly on two aspects of the theory.

Does Meaning Determine Reference?

Oddly, we now have a theory that is powerful enough to satisfy the criteria of adequacy for both meaning and truth, yet the theory is quite weak. So far as I can see, one cannot prove in this theory that:

If α and β have the same meaning, then they have the same reference.

Briefly, I think that the idea that meaning determines reference can be well justified, but for lack of time I will not go into the justification here. I merely assume it by adopting an axiom for *atomic* elements:

$$\text{If } \text{mng}\{\alpha\} = \text{mng}\{\beta\} \text{ then } \text{ref}\{\alpha\} = \text{ref}\{\beta\}.$$

We can then prove that if two sentences have the same meanings, then one is true if and only if the other is true. We will use this principle below.

Compositionality

A second issue concerns the compositionality of meanings. Here, the theory is also weak; we cannot prove that:

If α and β have the same meaning, then you can substitute one for the other without changing the meaning of the whole.

This is because no way has been given to calculate the meaning of the whole sentence from the meanings of its atomic parts. It *looks* as if we have done this, because there is a kind of system behind our atomic clauses, giving an appearance, for example, that the meaning of **white** contributes to the meaning of **Agatha believes that snow is white**. But that is not so. The theory determines the meaning of this whole sentence directly from the meaning of $\wedge\text{snow}$, not from the meaning of **snow**. And the meaning of $\wedge\text{snow}$ is, so far as we have specified, completely unrelated to the meaning of **snow**. For these are given independently of one another in the atomic axioms. You do not give the meaning of **snow** and then figure out the meaning of $\wedge\text{snow}$ from that information; it is simply specified independently. For example, we specify:

$$\begin{array}{ll} \text{mng}\{\mathbf{a}\} = \wedge\mathbf{a} & \text{mng}\{\mathbf{b}\} = \wedge\mathbf{b} \\ \text{mng}\{\wedge\mathbf{a}\} = \wedge\wedge\mathbf{a} & \text{mng}\{\wedge\mathbf{b}\} = \wedge\wedge\mathbf{b} \end{array}$$

If we suppose that **a** and **b** have been given the same meaning, it does not follow that we have specified the meanings of $\wedge\mathbf{a}$ and $\wedge\mathbf{b}$ to be the same.⁴ So if α and β have the same meaning, substituting one for the other within a caret can change the meaning of the whole. The theory is thus not a compositional one. This is undesirable.

2. SIMPLIFYING THE HIERARCHY OF SENSES

We have a theory of truth and meaning, but not a finitely stated theory. This is because we have an infinite number of axioms for each atomic word. This makes some people uncomfortable. More important, as noted, the theory is not compositional. It would be nice to have a compositional theory, and it would be nice to be rid of the complexity

4. We *are* required to specify each of the caretted terms as determining the meanings of the uncaredted terms, since meaning determines reference, but that doesn't mean that they have to be the same.

of the hierarchy of levels of meanings. I used to think that there was a way to accomplish this with ease. We just need some kind of relation that will make the meaning of $\hat{\text{is white}}$ a function of the meaning of is white . But what relation? The most elegant option here would be to let the relation be *identity*. That is, we state both the meaning and the reference of is white , and then identify the meaning of $\hat{\text{is white}}$ with that of is white . Since $\hat{\text{is white}}$ and is white never occur in the same contexts, they will never be confused with one another regarding their meaning, and no trouble will result. (That's the idea, anyway.)

We can accomplish this by adding as an axiom:

$$\text{mng}\{\hat{\alpha}\} = \text{mng}\{\alpha\}$$

We then drop all of the new atomic clauses that we added. Our axiom lets us replace any occurrence of nested carets by a single one:

$$\text{mng}\{\hat{\hat{\alpha}}\} = \text{mng}\{\hat{\alpha}\} = \text{mng}\{\alpha\}$$

All of our earlier proofs can be restated and proved using this assumption, and so we have eliminated the hierarchy and we have a simple compositional theory that meets the criteria of adequacy for both truth and meaning.

Burge's Argument about the Hierarchy of Senses

In an important but not widely studied paper,⁵ Tyler Burge argues that certain principles of Fregean semantics, if incorporated into a theory of truth, entail the infamous *Fregean hierarchy of senses*: the doctrine that each word must have associated with it not just one sense but an infinite number of them, because every time an indirect context is itself re-embedded in another indirect context, the contained words acquire *new* senses. For example, the underlined words in

Snow is white

express first-level senses and refer to non-senses; the same words in

Mary believes that snow is white

express second-level senses and refer to first-level senses; and the same words in

Agatha thinks that Mary believes that snow is white

express third-level senses and refer to second-level senses; and so on. Burge argues that if we take a Fregean approach, we will inevitably be stuck with such a hierarchy.

I am not sure if the principles that Burge cites are all in Frege, but they *are* in the semantical theory just developed. Not exactly the same

5. Tyler Burge, "Frege and the Hierarchy," *Synthese* 40 (1979): 304–323.

principles, but ones close enough that a version of Burge's argument can be formulated using our semantical rules as they stand. This will show that our rules, which at present eliminate such a hierarchy, are inconsistent (inconsistent with obvious data).

Burge's own argument deals with the truth-value of a double belief context (*Agatha believes that Mary believes that . . .*), but in our theory it is simpler to deal directly with the *meaning* of a sentence containing a *singly* embedded indirect context. So consider the meaning of **Mary believes that snow is white**. We begin by proving that this has the same meaning as **Mary believes snow is white**, where **snow is white** does *not* occur in an indirect context. The argument goes as follows, using 'S' for 'snow is white':

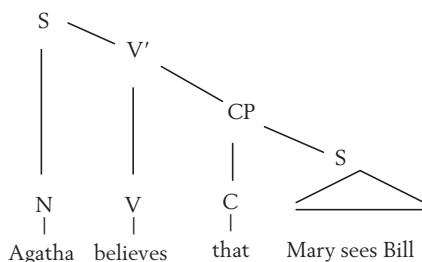
$$\begin{aligned} \text{mng}\{\mathbf{mB}^{\wedge}[\mathbf{S}]\} &= \text{mng}\{\mathbf{B}\}(\text{mng}\{\mathbf{m}\}, \text{mng}\{\mathbf{S}\}) && \text{PRED-ARG} \\ &= \text{mng}\{\mathbf{B}\}(\text{mng}\{\mathbf{m}\}, \text{mng}\{\mathbf{S}\}) && \text{E-MEANING} \\ &= \text{mng}\{\mathbf{mB} \mathbf{S}\} && \text{PRED-ARG} \end{aligned}$$

This shows that $\mathbf{mB}^{\wedge}[\mathbf{S}]$ and $\mathbf{mB} \mathbf{S}$ have the same meaning. But two sentences with the same meaning are alike in truth-value. So we can infer:

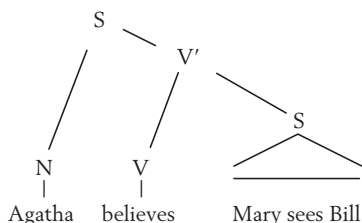
$\mathbf{mB}^{\wedge}[\mathbf{S}]$ is true iff $\mathbf{mB} \mathbf{S}$ is true.

This identifies the truth of a belief sentence (on the left) containing a sentence in an indirect context, with a belief sentence (on the right) with the same sentence *not* in an indirect context. In the semantics given earlier, sentences that are not embedded in that-clauses are interchangeable when they have the same truth-value without affecting the truth-value of the whole. So we have proved that sentences with the same truth-value are interchangeable within belief contexts! The theory thus contradicts the basic assumptions about data that motivated it. Disaster.

The oddity of this proof (as Burge himself notes) is that the verb 'believes' ends up combining with a sentence not in a that-clause. And the whole sentence that results is not *grammatical*. A normal sentence with a that-clause has a structure like this:



But Burge's argument also deals with the semantics of a sentence with a *different* structure:



This is not the grammatical structure of any sentence of English. In English, the word 'that' need not be present, but if it is absent we interpret the sentence as if it were present; we attribute to it the former structure, not the latter. This suggests a simple way to avoid Burge's problem: Limit our semantics to well-formed sentences, and the problem does not arise.⁶

For a long time I thought this was a good thing to do. But I have always had reservations. Couldn't there be a language like English, but in which sentences *can* follow 'believes' without being embedded in a complement clause? Indeed, can we be certain that English is not like this?

Worse, the problem goes deeper. It reappears in another guise. It is easiest to see the difficulty in principle if we ignore English for a moment and consider bare identity statements. Here is an adaptation of Burge's strategy, applied to the metalanguage itself:

Suppose that $\hat{a} = \hat{b}$.

Then $\hat{a} = (a, \hat{a}b) = \hat{a} = (\hat{a}a, \hat{a}b)$	COMP
$= \hat{a} = (\hat{a}a, \hat{a}b)$	SUBST IDEN
$= \hat{a} = (a, b)$	COMP

But for sentences, meaning determines truth-value. So what we just proved entails (putting the identity sign between its arguments):

$$a = \hat{a}b \leftrightarrow a = b$$

But in the special case where a and b are the same, this is:

$$a = \hat{a}a \leftrightarrow a = a$$

We have thus proved that each named thing is identical to the meaning of any name that refers to it. Disaster re-arises even without looking at ungrammatical sentences.⁷ There are ways to try to get around this, but

6. Recall that we were confident about identifying the meanings of S with $\hat{S}$ because they could never appear in the same context. Burge's argument assumes that they can.

7. Focusing on well-formed sentences of the object language is irrelevant, since this argument proceeds entirely in the metalanguage, using only identity sentences of the sort that we have taken for granted are well formed throughout our theorizing.

none that I now have any confidence in. This motivates us to seek a different way to avoid the hierarchy, or to live with it.

3. A ROAD BACK

It is apparent that that-clauses themselves have meaning. When I reflect on the meanings of that-clauses, it seems obvious that the meaning of the that-clause itself is *determined* by the meaning of the sentence it contains. In interpreting the that-clause, all you have to go on is the word ‘that’ and the sentence following it. From these ingredients alone we understand the that-clause. Since the meaning of ‘that’ is constant, the determining is done by the meaning of the sentence in the clause. But the meaning of the contained sentence is what the *whole* that-clause refers to. So the meanings of that-clauses are determined by their referents. For that-clauses, there is a road back from reference to meaning.

How can this be? Well, in the essay “Quantifying In,” by someone whose name escapes me at the moment, the author considers how to analyze *de re* readings of sentences in a Fregean framework. In a digression, he discusses what he calls ‘standard names’. Examples of standard names are numerals. The numeral ‘18’ designates the same number as ‘ 9×2 ’, but ‘18’ is a *standard* name of that number, whereas ‘ 9×2 ’ is not.

He also suggested that quotation mark names are standard names of the expressions that occur between the marks; the name:

‘Socrates is wise’

including the quotation marks, refers to the sentence contained between the quotes. If that is so, it seems equally natural to see a that-clause like

that Socrates is wise

as a standard name of the proposition that is expressed by the sentence following the ‘that’. This idea requires that there be a functional relationship between meanings of sentences—propositions—and meanings of the that-clauses that standardly refer to those propositions. I do not know how to define this relation in familiar terms, so I will just introduce a symbol for it; I will use an up-arrow ‘ $\uparrow$ ’ to stand for the function that maps each proposition to the unique meaning of its standard name—to the meaning of any that-clause which refers to it.⁸ To characterize it, we introduce an appropriate new axiom to our semantic theory, the principle:

UP-ARROW $\text{mng}\{\wedge S\} = \uparrow(\text{mng}\{S\})$

8. I am adding the up-arrow to the metalanguage, not to the object language. I see nothing in the object language that corresponds to it.

In the metalanguage we assume the new logical principle:

$$\text{UP-ARROW } \hat{\alpha} = \uparrow(\hat{\alpha})$$

With these principles in hand, we can now truncate our infinite set of atomic axioms, leaving only the genuinely atomic ones, such as:

$$\begin{aligned} \text{ref}\{\text{is white}\} &= \text{is white} \\ \text{mng}\{\text{is white}\} &= \hat{\text{is white}} \end{aligned}$$

The rest may be recovered from the general principles we just adopted.

We then have a simple theory again, and we can work through the proofs that we had before, and they work fine. Example:

$$\text{mng}\{\text{Mary Believe } \hat{[\text{Wise Socrates}]}\}$$

$$\begin{aligned} \text{PRED-ARG-M} &= \text{mng}\{\text{Believe}\}(\text{mng}\{\text{Mary}\}, \text{mng}\{\hat{[\text{Wise Socrates}]}\}) \\ \text{ATOMIC} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \text{mng}\{\hat{[\text{Wise Socrates}]}\}) \\ \text{COMP} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \text{mng}\{\hat{[\text{Wise Socrates}]}\}) \\ \text{M-IDEN} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \text{ref}\{\hat{[\text{Wise Socrates}]}\}) \\ \text{COMP} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \text{ref}\{\hat{\hat{\text{Wise}}} \hat{\hat{\text{Socrates}}}\}) \\ \text{REF-FUNC} &= \text{Believe}(\hat{\text{Mary}}, \text{ref}\{\hat{\hat{\text{Wise}}}\}(\text{ref}\{\hat{\hat{\text{Socrates}}}\})) \\ \text{M-IDEN} &= \text{Believe}(\hat{\text{Mary}}, \text{mng}\{\hat{\text{Wise}}\}(\text{mng}\{\hat{\text{Socrates}}\})) \\ \text{PRED-ARG-M} &= \text{Believe}(\hat{\text{Mary}}, \text{mng}\{\hat{\text{Wise}} \hat{\text{Socrates}}\}) \\ \text{COMP} &= \text{Believe}(\hat{\text{Mary}}, \text{mng}\{\hat{[\text{Wise Socrates}]}\}) \\ \text{UP-ARROW} &= \text{Believe}(\hat{\text{Mary}}, \uparrow(\text{mng}\{\text{Wise Socrates}\})) \\ &\quad [\text{but we proved earlier that } \text{mng}\{\text{Wise Socrates}\} = [\text{Wise}(\text{Socrates})].] \\ \text{so:} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \uparrow(\hat{[\text{Wise}(\text{Socrates})]})) \\ \text{UP-ARROW} &= \hat{\text{Believe}}(\hat{\text{Mary}}, \hat{\hat{[\text{Wise}(\text{Socrates})]}}) \\ \text{COMP} &= \hat{[\text{Believe}(\text{Mary}, \hat{[\text{Wise}(\text{Socrates})]})]}. \end{aligned}$$

We can get the same proofs as before, and the Burge problem does not arise because we never *identify* the meaning of an expression with the meaning of that expression preceded by a meaning-caret. In fact, we have not eliminated a hierarchy of senses at all; we merely view it as a kind of epiphenomenon of the simple part at the basis of the hierarchy.

An Application

Russell says that we should always have some puzzles in our toolbox with which to test semantic theories. One of my favorite puzzles is how to account for quantification over propositions into indirect contexts. An example is this argument, which seems to be valid:

George believes whatever Agatha thinks Maria believes
 Agatha thinks that Maria believes that snow is white
 $\therefore$ George believes that snow is white

Or, subscripting the believers to reduce clutter:

$$\begin{aligned} &\forall x [B_a^\wedge[B_mx] \rightarrow B_gx] \\ &B_a^\wedge[B_m^\wedge[Ws]] \\ &\therefore B_g^\wedge[Ws] \end{aligned}$$

How can we account for this validity? If we instantiate the variable ‘x’ with the meaning name ‘ $^\wedge[Ws]$ ’ then the argument goes through. Because the first premise instantiates to:

$$B_a^\wedge[B_m^\wedge[Ws]] \rightarrow B_g^\wedge[Ws]$$

and modus ponens does the rest. But why should we think that our semantic theory says that the variable ‘x’ is in a position where we *can* instantiate it? Its quantifier quantifies into an indirect context—in symbols, it binds a variable within a context governed by a meaning-caret, and it is not obvious how this is to work.

Here is an idea. If our previous rules were to apply to variables just as to constants, the antecedent of the conditional:

$$B_a^\wedge[B_mx]$$

would be equivalent to:

$$B_a^\wedge B_m(^{\wedge}x)$$

But what does ‘ $^\wedge x$ ’ mean when ‘x’ is a variable? That is, what is the value of ‘ $^\wedge x$ ’ when we assign an object to the variable ‘x’? Or, what does ‘ $^\wedge x$ ’ refer to relative to an assignment to ‘x’?

To answer this, let us return to “Quantifying In.” The goal of “Quantifying In” was to explain *de re* occurrences of things. The reason the author discussed standard names was that using a standard name in an indirect context gives you the *effect* of having a name that occurs *de re*.

(If we reflect on our treatment of that-clauses, this means that a that-clause always occurs *de re*. A test for this is to ask whether a sentence with a that-clause can have two readings, a *de re* one and a *de dicto* one. George Bealer has pointed out that the data suggests that this does not happen with that-clauses.)⁹

I would like to extend Kaplan’s view in a way that he either disavowed or did not discuss, and suggest that in a sense *every* object can have something like a standard name; it is a variable with that object assigned to the variable.

9. The argument here is parallel to the difference between using an ordinary definite description:

Necessarily, the number of planets is odd

and using a standard name:

Necessarily, 9 is odd.

I propose to identify the reference of a caret variable relative to an assignment of object o to that variable with the up-arrow of o . That is, if 'x' is a variable, then:

$$\text{ref}_\sigma\{\wedge x\} = \uparrow(\text{ref}_\sigma\{x\})$$

where " $\text{ref}_\sigma\{\alpha\}$ " is the reference of α relative to the assignment σ of things to the variables of the language. Then in the metalanguage we should assert that in any context, a formula with a variable preceded by a caret is equivalent to the same formula with the variable instead preceded by the up-arrow. This gives us exactly what we need in the above case. The first premise in the argument is:

$$\forall x [B_a \wedge [B_m(x)] \rightarrow B_g x]$$

This is equivalent by earlier rules to:

$$\forall x [B_a \wedge B_m(\wedge x) \rightarrow B_g x]$$

which we have said would be:

$$\forall x [B_a \wedge B_m \uparrow(x) \rightarrow B_g x]$$

The variable 'x' is here in an extensional position, and so we can freely instantiate it with the meaning name ' $\wedge[Ws]$ ', to get:

$$\forall x [B_a \wedge B_m \uparrow(\wedge[Ws])] \rightarrow B_g \wedge[Ws]$$

Now we can convert the interior part of the antecedent (which is in an extensional position):

$$\uparrow(\wedge[Ws])$$

to:

$$\wedge\wedge[Ws]$$

(using UP-ARROW), giving us:

$$\forall x [B_a \wedge B_m \wedge\wedge[Ws] \rightarrow B_g \wedge[Ws]]$$

which, by previous rules, converts to:

$$\forall x [B_a \wedge [B_m \wedge\wedge[Ws]] \rightarrow B_g \wedge[Ws]]$$

And this is where modus ponens does the rest.

In summary, the theory that results from all of this meets the criteria of adequacy for both truth and meaning, it is compositional with respect to meaning, and it is semantically innocent: No word or phrase ever changes its reference or its meaning in any context.

CHAPTER 5

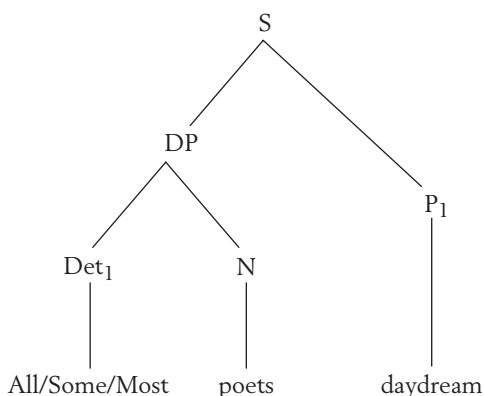
Some Logical Properties of Natural Language Quantifiers

Edward L. Keenan

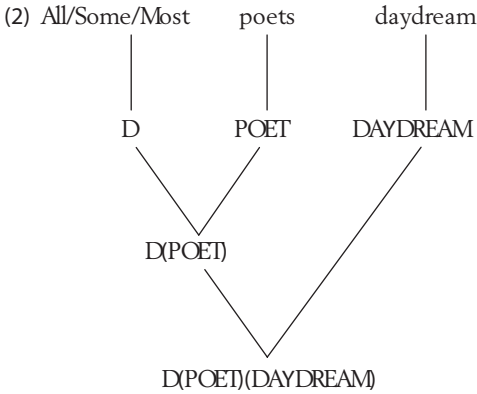
1. BACKGROUND

In English, quantifiers may occur adverbially (*John often/rarely/never visits museums in summer*), but they occur most productively as one-place determiners (Det_1 s), which combine with Nouns (Ns) to form *Determiner Phrases* (DPs). These in turn combine with one-place predicates, P_1 s, to form Ss (Sentences):

(1)



Semantically we interpret Ns and P_1 s as properties of objects, extensionally represented as subsets of whatever domain E of objects is under discussion. We interpret Ss standardly as True or False and, initially, we interpret DPs as functions called *generalized quantifiers* (GQs) mapping P_1 denotations to $S (= P_0)$ denotations. Det_1 s denote functions mapping sets to GQs:



Ss like those in (2) are interpreted as whatever truth-value the GQ denoted by the subject DP maps the set denoted by the P_1 to. For example, ALL(POET) maps a set B (such as DAYDREAM) to True iff each object in the POET set is also in B, that is, iff $POET \subseteq B$. And the Det_1 denotation ALL maps each subset A of E to ALL(A), that generalized quantifier that maps a set B to True iff $A \subseteq B$. Similarly SOME(POET) maps B to True iff there is an object in both POET and B, that is, iff $POET \cap B \neq \emptyset$. In general SOME maps a set A to SOME(A), that GQ which maps a set B to True iff $A \cap B \neq \emptyset$. Some further examples:

(3) a. NO(A)(B)	= True iff $A \cap B = \emptyset$
b. (ALL BUT SIX)(A)(B)	= True iff $ A - B = 6$
c. MOST(A)(B)	= True iff $2 \bullet A \cap B > A $
d. (AT MOST TWO OUT OF THREE)(A)(B)	= True iff $3 \bullet A \cap B \leq 2 \bullet A $
e. (THE TEN)(A)(B)	= True iff $ A = 10$ and $A \subseteq B$
f. NEITHER(A)(B)	= True iff $ A = 2$ and $A \cap B = \emptyset$
g. (AT MOST SIX OF THE TEN)(A)(B)	= True iff $ A = 10$ and $ A \cap B \leq 6$

In writing D(A)(B) above we call A the *noun* argument and B the *predicate* one.

So here Det_1 s are semantically interpreted. CL (Classical Logic) introduces them syncategorematically and interprets $\forall x\Phi$ and $\exists x\Phi$ in terms of Φ , but does not directly interpret $\forall x$ or $\exists x$. By assigning a type to Dets and DPs we make them objects of study, and so are in a better position than classically to discover generalizations about their behavior in natural language. Note too that we have separated quantification from variable binding, notions that are conceptually distinct and expressed by distinct means in English: quantification by Dets and adverbs, binding by pronouns (*Every poet admires himself / Most professors think they're clever*) or “gaps” (*every student who John met _ at the party*). CL, by contrast, expresses quantification and binding in the same operators: In $\forall x\Phi$ and

$\exists x\Phi$, $\forall x$ and $\exists x$ bind free occurrences of x in Φ in the same way but differ in quantificational force.

It is easy to modify CL to separate quantification and variable binding: Use the lambda operator to convert a formula Φ to a P_1 , $\lambda x\Phi$, that denotes the set of objects b such that Φ is True when x is set to denote b . Then $\forall$ and $\exists$ are introduced as functions mapping P_1 denotations to S denotations: $\forall$ maps B to True iff $B = E$, the domain of the model, and $\exists$ maps B to True iff $B \neq \Phi$. One verifies that $(4a, a')$ are logically equivalent, as are $(4b, b')$.

- (4) a. $\forall x(\text{Poet}(x) \rightarrow \text{Admire}(x, x))$ a'. $\forall(\lambda x(\text{Poet}(x) \rightarrow \text{Admire}(x, x)))$
 b. $\exists x(\text{Poet}(x) \& \text{Admire}(x, x))$ b'. $\exists(\lambda x(\text{Poet}(x) \& \text{Admire}(x, x)))$

We may now compare quantification in CL with that in English. The obvious difference is that $\forall$ and $\exists$ are unary, taking simply a single set as argument. But the English quantifiers in (3) are binary, taking two sets as arguments. It is the noun property that is “new.”

2. LOCAL PROPERTIES OF ENGLISH QUANTIFIERS

2.1. The Noun Property

The noun property functions to limit the objects we need consider in evaluating $D(A)(B)$. In well studied first-order languages, such as that of elementary arithmetic or set theory, there is an understood domain of objects we use the language to talk about and which the variables range over. But natural languages are used in a great variety of contexts whose understood domains vary. The role of the Ns that Dets combine with is to provide “on line” the domain intended for purposes of interpreting predication involving that DP.

If we limited our study to English versions of classical quantifiers, *some* and *all* and their paraphrases, the use of Ns to adjust the domain to the communicative needs of the moment would appear to be logically inessential (though perhaps convenient), as Frege realized (Westerståhl 1989). The standard translations of (5a,b) as (5a',b') respectively show how to paraphrase Ss in which the domain of quantification is restricted by ones in which it is not.

- (5) a. All poets daydream a'. $\forall x(\text{Poet}(x) \rightarrow \text{Daydream}(x))$
 b. Some poets daydream b'. $\exists x(\text{Poet}(x) \& \text{Daydream}(x))$

But it turns out that English presents Dets that do not admit of such *sortal reduction*. That is, the noun restriction on the domain of quantification cannot be eliminated in favor of quantifying over the entire domain and compensating in the predicate by using some Boolean compound of the original noun and predicate properties. Thus quantification in English is essentially binary (at least). Formally, using D_E as a variable ranging over possible Det denotations relative to E , we define:

Def 1 D_E is *sortally reducible* iff for all properties A,B $D_E(A)(B) = D_E(E)(h(A,B))$, where 'h' is a fixed two-place Boolean function.

"Boolean functions" are ones definable solely in terms of the Boolean connectives *and*, *or*, *not*, and so forth. This just means that $h(A,B)$ above is defined in terms of set intersection, union, and complement. Classical *all* and *some* are sortally reducible:

- (6) a. $SOME(A)(B) = SOME(E)(A \cap B)$
 "Some As are Bs" $\equiv$ "Some entities are both As and Bs"
 b. $ALL(A)(B) = ALL(E)(\neg A \cup B)$
 "All As are Bs" $\equiv$ "All entities are either not As or else Bs"

(Using the abbreviation $X \rightarrow Y$ for $\neg X \cup Y$ the right-hand side of (6b) reads more naturally as "All entities are such that if they are As then they are Bs"). And we have:

Claim 1 English presents Dets that are not sortally reducible

Theorem 7 below characterizes the sortally reducible Det_1 s. It builds on a classification of Dets given below plus one property all Dets satisfy. So here we just give an example.

- (7) *most*, in the sense of "more than half" as in (3c), is not sortally reducible

Let us see that the reductions in (6) do not work. (6a) clearly does not, since *Most students are vegetarians* is not logically equivalent to *Most entities are both students and vegetarians*: Given a 100-element domain all of whose members are vegetarians and only 15 of whom are students, the first sentence is true and the second false. In the second case, the one speakers tend to try when seeking a reduction, compare (8a) and (8b):

- (8) a. Most students are vegetarians
 b. Most entities are such that if they are students then they are vegetarians

To see that (8b) does not entail (8a) consider a model with 100 entities; just 15 are students, none are vegetarians. Then (8a) is clearly false, since none of the students are vegetarians. But (8b) is true, since the *if-then* clause is satisfied by 85 of the 100 entities.

An Historical Aside

Russell and Frege had somewhat different conceptions of quantifiers, with Frege's closer to ours, as pointed out in Westerståhl (1989):

"It must here be remarked that the words 'all', 'any', 'no', 'some' are prefixed to concept-words. In universal and particular affirmative and negative sentences, we are expressing relations between concepts; we use these words to indicate the special kind of relation." (Frege 1892: 48)

This passage continues with the example *all mammals are land dwellers*, showing that *concept-words* include Ns and P_1 s in general. But Russell

(1908), in interpreting *All men are mortal*, destroys the constituent *all men* and treats the quantifier as binding a variable whose range properly includes the set of men:

“‘All men’ is a denoting phrase; . . . denoting phrases never have any meaning in isolation, but only enter as constituents into the verbal expression of *propositions which contain no constituent corresponding to the denoting phrases in question*. . . . Hence it is impossible that these propositions should acquire their meaning through the denoting phrases.” (Russell 1908: 70; emphasis mine)

“What is meant by ‘all men are mortal’ may be more explicitly stated in some such form as ‘it is always true that if x is a man, x is mortal’ [Russell 1908:71]. . . . it follows that the function ‘if x is a man, x is mortal’ must have a certain *range of significance*, which falls short of all imaginable values of x , though it exceeds the values which are men. The restriction on x is therefore a restriction to the range of significance of the function ‘if x is a man, x is mortal’ [72]. . . . For example, let us attempt to restrict the variable to *men*, and assert that, subject to this restriction, ‘ x is mortal’ is always true. Then what is always true is that if x is a man, x is mortal; and this hypothetical is true even when x is not a man. Thus a variable can never be restricted within a certain range if the propositional function in which the variable occurs remains significant when the variable is outside that range.” (73)

Russell’s view, which held sway for a century or so, makes it impossible to treat the many non-sortally reducible quantifiers in natural language discussed in section 3.

2.1.1. Conservativity

Generally in English when a generalized quantifier $D(A)$ applies to a property B , its value depends only on those elements of B that lie in A . So $D(A)$ yields the same value when applied to any other B' that has the same intersection with A as B does, a condition known as *conservativity*: Writing P_E for the set of subsets of E , and $[X \rightarrow Y]$ for the set of functions from X into Y , we define *conservativity* as an invariance condition, Def 2. The more usual definition is now a theorem, Proposition 1:

Def 2 (CONS) D from P_E into $[P_E \rightarrow X]$, X any set, is *conservative* iff for all $A, B, B' \subseteq E$, $D(A)(B) = D(A)(B')$ whenever $A \cap B = A \cap B'$

Prop 1 A function D from P_E to $[P_E \rightarrow X]$ is conservative iff for all $A, B \subseteq E$, $DAB = DAA \cap B$.

Using Prop 1 verify the conservativity of Det_{1s} by verifying logical equivalences like:

(9) a. All swans are white	a'. All swans are swans and are white
b. Most students read the <i>Times</i>	b'. Most students are students who read the <i>Times</i>
c. No dogs bark	c'. No dogs are dogs that bark

- d. John's students work hard d'. John's students are both students and work hard
 e. Which countries have no taxes? e'. Which countries are countries with no taxes?

The choice of denotation set for S_s is immaterial in deciding whether the Det_1 s are conservative. Def 2 just requires that $D(A)$ yield the *same* value at B 's that have the same intersection with A . We might choose declarative S_s to denote functions from possible worlds to truth-values without changing the Det_1 s classed as conservative.

This added generality is helpful in interpreting interrogative Dets, such as *Which?*, *Whose?*, and *How many?* The interpretation of questions is debatable (see Groenendijk and Stokhof 1997), but it seems clear that interrogative Dets are conservative, as we judge that the (x, x') pairs of questions below request the same information with x' redundant.

- (10) a. How many athletes use steroids?
 a'. How many athletes are athletes who use steroids?
 b. Which students read the *Times*?
 b'. Which students are students and read the *Times*?
 c. Whose two books are on the table?
 c'. Whose two books are books that are on the table?

If "five" is a true answer to (10a), then it is a true answer to (10a') and conversely. If "just John and Bill" answers (10b), then it answers (10b') and conversely, and so on. So even if we are not sure what sorts of semantic objects questions denote, we have reasonable judgments that (11) holds (where WH ranges over HOW MANY?, WHICH?, WHOSE TWO?, etc.).

- (11) For all properties A, B, B' if $A \cap B = A \cap B'$ then $\text{WH}(A)(B)? = \text{WH}(A)(B')?$

A further generalization (Keenan and Stavi 1986) in the statement of conservativity would show that our choice of an extensional set theoretic formalism is unnecessarily concrete. The only important property of the definition is that the arguments of D lie in a partially ordered set in which all elements x, y have a greatest lower bound (glb) (noted $x \wedge y$). Power sets are such sets in which the glb of two elements is their intersection. So a more general statement of Def 2 would replace $A \cap B$ with $A \wedge B$, and so on. We do not pursue this generalization here though it is conceptually helpful in representing Boolean compounds of non-count nouns, as in *There was neither joy nor sadness in his voice*.

The equivalences in (9) are felt as trivial because the predicate repeats information given by the noun. Nonetheless, CONS is a very strong constraint, as attested by the following cardinality facts, writing GQ_E for $[P_E \rightarrow \{\text{True}, \text{False}\}]$, the set of GQs over E :

Theorem 2 The cardinality of GQ_E is 2^k , for $k = 2^{|E|}$. The number of logically possible Det_1 denotations is the cardinality of $[P_E \rightarrow \text{GQ}_E]$, namely, 2^k , for

$k = 4^{|E|}$. The number of these functions which are conservative is just 2^k for $k = 3^{|E|}$ (Keenan and Stavi 1986).

Thus in a two-element domain there are $2^{16} = 65,536$ maps from properties to GQs. Just $2^9 = 512$ are conservative! Functions like H (the Härtig quantifier) are not conservative: choosing A and B to be disjoint nonempty sets of the same size, $|A| = |B|$ but $|A| \neq |A \cap B|$.

$$(12) H(A)(B) = \text{True iff } |A| = |B|$$

2.2. Three Semantic Classes of Dets

In deciding what value to assign to a pair A, B of sets, a conservative D (We often omit the subscript on D_E) may be sensitive to properties of A , $A \cap B$, or $A - B$. Here we classify Dets according to how much of these three sets they are sensitive to.

2.2.1. Intersective Dets

Intersective Dets are only sensitive to $A \cap B$, the A s that are B s:

Def 3 D is *intersective* iff for all A, A', B, B' , $D(A)(B) = D(A')(B')$ if $A \cap B = A' \cap B'$.

Some is intersective, since if the A s that are B s are the same as the X s that are Y s then *Some As are Bs* iff *Some Xs are Ys*. For this reason we sometimes call intersective Dets *generalized existential* Dets. Props 3a,b help in establishing the intersectivity of a Det.

Prop 3a D is intersective iff for all properties A, B $DAB = D(A \cap B)(E)$.

We may read " $D(A \cap B)(E)$ " as " D As that are Bs exist".

Prop 3b D is intersective iff for all A, B $DAB = D(E)(A \cap B)$.

We may read $D(E)(A \cap B)$ as " D entities are both As and Bs".

Since the (x, x') pairs in (13) are logically equivalent the Dets in question are intersective. Similarly, we see that the Dets in (14) are intersective (though some cases are subtle).

(13) a. Some students are vegetarians	a'. Some students who are vegetarians exist
b. Exactly five boys are in the room	b'. Exactly five entities are boys in the room
c. More male than female pupils cried	c'. There exist more male than female pupils who cried

(14) some, a/an, no, several, more than six, at least six, exactly six, fewer than six, at most six, between six and ten, just finitely many, infinitely many, about a hundred, a couple of dozen, practically no, nearly twenty, approximately twenty, not more than ten, at least two and not more than ten, either fewer than five or else more than twenty, that many, How many?, Which?, more male than female, just as many male as female, no ... but John

Most of the Dets in (14) are not merely intersective; they are *cardinal*, meaning that $D(A)(B)$ is determined just by *how many* elements are in $A \cap B$. We do not have to know precisely which entities they are. So the truth of *Between five and ten boys are in the class* is decided just by giving $|\text{BOY} \cap \text{IN THE CLASS}|$, the number of boys in the class:

Def 4 A possible Det_1 denotation D is *cardinal* iff for all sets A, B, A', B' if $|A \cap B| = |A' \cap B'|$ then $DAB = DA'B'$

Being cardinal is a very special case of being intersective.¹ An advantage of Def 4 is that it makes clear that vague Dets such as *about fifty*, *nearly a hundred*, *a few more than a hundred*, and so on are cardinal: If the number of sparrows on my clothesline is the same as the number of coins in that fountain then about fifty sparrows are on my clothesline iff about fifty coins are in that fountain (even though we may disagree on precisely how many coins are to count as “about fifty”). Also, under these conditions, *How many sparrows are on your clothesline?* and *How many coins are in that fountain?* must have the same true answers, so the interrogative Det *How many?* is cardinal.

The intersectivity condition is extremely strong: Of the 512 conservative functions in a world of just two individuals, only 16 are intersective! To see this, observe first that Boolean compounds of Dets are interpreted pointwise, as in (15), given formally in (16).

- (15) (Some but not all) cats are black $\equiv$ (Some cats but (not all) cats) are black
 $\equiv$ (Some cats but (not (all cats))) are black
 $\equiv$ Some cats are black but (not (all cats)) are black
 $\equiv$ Some cats are black but not (all cats are black)

- (16) a. $(D \vee D')(A)(B) =_{\text{def}} (D(A) \vee D'(A))(B) = D(A)(B) \vee D'(A)(B)$.
 b. $(D \wedge D')(A)(B) =_{\text{def}} (D(A) \wedge D'(A))(B) = D(A)(B) \wedge D'(A)(B)$.
 c. $(\neg D)(A)(B) =_{\text{def}} \neg(D(A))(B) = \neg(D(A)(B))$.

1. The cardinal functions are a very proper subset of the intersective ones (when $|E| > 1$). Writing INT_1 (CARD_1) for the set of intersective (cardinal) functions from P_E to GQ_E , we have from Keenan and Moss (1985):

Prop 4 For all E , $|\text{CARD}_1| = 2^{|E| + 1}$ and $|\text{INT}_1| = 2^k$, for $k = 2^{|E|}$.

Proof Sketch For the first claim, show that CARD_1 is a complete atomic Boolean algebra pointwise whose atoms are $\{D_k | 0 \leq k \leq |E|\}$, where $D_k(A)(B) = 1$ iff $|A \cap B| = k$. For the second, show that the map h sending each intersective function D to $D(E)$ is an isomorphism from INT_1 to $\text{GQ} = [P_E \rightarrow \{\text{True}, \text{False}\}]$, regarded as a Boolean algebra pointwise.

So in a model with $|E| = 4$, there are $2^5 = 32$ possible denotations for cardinal Det_1 s but $2^{16} = 65,536$ possible denotations for the full set of intersective Det_1 s!

Theorem 4a The set of intersective functions is closed under pointwise $\wedge$, $\vee$, and $\neg$, whence this set is a Boolean algebra (in fact a Boolean subalgebra of the set of conservative functions, also closed under pointwise $\wedge$, $\vee$, and $\neg$).

Theorem 4b The function h mapping each intersective Det_1 function D to the generalized quantifier $D(E)$ is a Boolean isomorphism (Keenan 1993).

From Theorem 4b we see that the intersective *Det some* corresponds to the GQ *some entity, between five and ten* to *between five and ten entities*, and so on. And as a corollary we have that the number of intersective functions is always the same as the number of GQs. On reflection this is not surprising. The value of an intersective D at properties A, B is decided by a property of a single set, $A \cap B$. So the intersective D 's are, in effect, just properties of sets.

2.2.2. Co-intersective (Generalized Universal) Dets

Co-intersective Dets just consider $A - B$, the A s that are not B s: Here first are some examples, followed by a proper definition:

(17) a.	$\text{ALL}(A)(B)$	= True iff $A - B = \emptyset$
b.	$(\text{ALL BUT SIX})(A)(B)$	= True iff $ A - B = 6$
c.	$(\text{ALL BUT FINITELY MANY})(A)(B)$	= True iff $A - B$ is finite
d.	$(\text{EVERY} \dots \text{BUT JOHN})(A)(B)$	= True iff $A - B = \{\text{John}\}$

Def 5 D is *co-intersective* iff it satisfies the following invariance condition:

$$DAB = DA'B' \text{ whenever } A - B = A' - B' \text{ all } A, A', B, B'$$

And we may define the *co-cardinal* Dets as those D satisfying $DAB = DA'B'$ whenever $|A - B| = |A' - B'|$. The Dets in (17a,b,c) are co-cardinal, that in (17d) is not.

The co-intersective Dets are structurally less diverse than the intersective ones. They are basically just the universal quantifier with exception phrases. *How many?* and *Which?* are intersective (with the first cardinal). But our best attempt at a co-intersective interrogative Det is *All but how many?* as in (18a), of dubious grammaticality. It means the same as the natural (18b) built from the intersective *How many?*

- (18) a. ?All but how many students came to the party?
 b. How many students didn't come to the party?

Note that the value of a co-intersective Det at properties A, B is decided by a property of a single set, $A - B$. So the number of co-intersective functions is the same as the number of GQs, the map h sending each co-intersective D to $D(E)$ is a Boolean isomorphism, noting that as with the intersective Dets, the co-intersective ones are closed under pointwise $\wedge$, $\vee$, and $\neg$.

These cardinality facts lead us to a somewhat surprising observation. Note first that the classical existential quantifier *some* (= *at least one*) and

other cardinality quantifiers (*at least n*, *fewer than n*, *between n and m*, etc.) all lie in the intersective class. Similarly, the universal type Dets, *all*, *all but n*, and so on all lie in the co-intersective class. Yet, in terms of denotations, these two classes account for only a very small percentage of the conservative functions. In a world of just two objects, there are only 16 intersective functions and 16 co-intersective ones,² but a total of 512 conservative functions. So the vast majority of conservative functions are neither intersective nor co-intersective.³

What are they? **Prop 5** gives an answer: Observe that nontrivial Boolean compounds of properly intersective with properly co-intersective Dets, as in (19a,b), are themselves neither intersective nor co-intersective.

- (19) a. Either all or else not more than two of the students will pass that exam
 b. Either just two or three or else all but a couple of students will pass that exam

Prop 5 (Keenan 1993) The set of conservative functions over E is precisely the Boolean closure of the intersective together with the co-intersective functions.

Thus from a Boolean perspective we can say that if you understand the (co)-intersective functions and you understand how to interpret Boolean compounds, you understand all conservative functions (despite the semantic clumsiness of the compounds themselves).

2.2.3. Proportionality Dets

Proportionality Dets (see Prop 6) are conservative but lie outside the (co)-intersective ones. They decide DAB according to the proportion of

2. The two trivial Det_1 functions **0**, which maps all A,B to False, and **1**, which maps all A,B to True, are both intersective and co-intersective, and are the only Det_1 functions with this property. Thus in a two-object world there are $(16 + 16) - 2 = 30$ functions that are either intersective or co-intersective, and thus $512 - 30 = 482$ that are neither. Thus the great majority of conservative functions are neither intersective nor co-intersective.

3. Certain other Dets behave somewhat like the co-intersective ones if, along with Barwise and Cooper (1981), we were to treat them as partial functions. Consider:

- i. $(\text{THE } n)(A)(B) = \text{True}$ iff $|A| = n$ and $\text{ALL}(A)(B)$
- ii. $(\text{JOHN's } n)(A)(B) = \text{True}$ iff $|A \cap \{b | \text{JOHN HAS } b\}| = n$ & $\text{ALL}(A \cap \{b | \text{JOHN HAS } b\})(B)$

On this semantics *the ten* and *John's ten* behave like co-intersective Dets when the cardinality condition on their first argument is satisfied. That condition removes them from the co-intersective class as we defined it. Barwise and Cooper would treat THE TEN as a partial function, one whose domain was the set of A with exactly ten members. But on that semantics as well *the ten* does not denote a co-intersective function. Consider A with cardinality 10 and A' with cardinality 5. Then $(\text{THE TEN})(A)(A) = \text{True}$ and $(\text{THE TEN})(A')(A')$ is undefined, so THE TEN fails to be co-intersective. But if we modified the definition of co-intersective to read $\text{DAB} = \text{DA'B'}$ when $A - B = A' - B'$ and both $A, A' \in \text{Dom}(D)$ then THE TEN, etc. would be co-intersective.

As that are Bs. So, they yield the same value at A,B as they do at A',B' when the proportion of As that are Bs is the same as the proportion of A's that are B's. Here are some examples:

- (20) a. Most poets daydream
 b. Seven out of ten sailors smoke Players
 c. Less than half the students here got an A on the exam
 d. At most 10 percent of the students passed the exam
 e. Between one-third and two-thirds of the students are vegetarians
 f. All but a tenth of the students are vegetarians
 g. Not one student in ten can answer that question

Def 6 D is *proportional* iff $DAB = DA'B'$ whenever $|A \cap B|/|A| = |A' \cap B'|/|A'|$ ⁴

Here are a few proportional Det denotations, just to make clear that they satisfy Def 6. We stipulate that they map A,B to False whenever A is empty or infinite (in which case the everyday notion of proportion does not make much sense). Otherwise:

- (21) a. (SEVEN OUT OF TEN)(A)(B) = True iff $10 \times |A \cap B| = 7 \times |A|$
 b. (AT MOST 10%)(A)(B) = True iff $10 \times |\bar{A} \cap B| \leq |A|$
 c. (ALL BUT A TENTH)(A)(B) = True iff $10 \times |A - B| = |A|$
 d. (NOT ONE...IN TEN)(A)(B) = True iff $10 \times |A \cap B| < |A|$

Note that apparently “co-proportional” Dets such as *all but a tenth*, which refer to the proportion of As that lack B, are covered by Def 6, since the proportion of As with B is uniquely determined by the proportion that lack B: *All but (at most) a tenth* means the same as *(At least) nine out of ten*, so (22a,b) are logically equivalent.

- (22) a. All but at most a tenth of the students read the *Times*
 b. At least nine out of ten students read the *Times*

Fractional and percentage Dets and ones of the form *n out of m*, $n \leq m$, and the discontinuous *less than/more than/exactly one...in ten* are proportional. They have been less studied than the generalized existential and universal ones. Interrogative members of this class seem more natural than for the universal ones but less than for the existential ones. Our best candidates are *What percentage of?* and *What proportion of?*, as in (23a) where they could be replaced by interrogative *Which?*, preserving grammaticality (but not proportionality)

- (23) a. What percentage / proportion of students read the *Times*?
 b. *Whatnth students read the *Times*?

(23a) asks for a proportion, but possibly does so compositionally using *What?* and the N *proportion of the students*. A more fully grammaticized

4. To avoid the case of division by zero we replace the definiens by $|A'| \cdot |A \cap B| = |A| \cdot |A' \cap B'|$.

interrogative proportional Det would be *Whatnth?* as in (23b), which means the same as (23a) but, alas, lacks existence.

Typically proportionality Dets are neither intersective nor co-intersective, but there are a few exceptions: *exactly zero percent* . . . means *no*, and *more than zero percent* . . . means *at least one*, so both are intersective; *a hundred percent* . . . means *all* and *less than a hundred percent* . . . means *not all*, so both are co-intersective. So the proportionality Dets have a nonempty intersection with the (co)-intersective Dets, but their precise relation is open.

There are yet other sorts of expressible Dets in English, such as (*two of the ten*), but the three classes so far discerned are all *logically coherent* in the sense that (Prop 6) they are closed under the formation of Boolean compounds (in *and*, *or*, *not*, *neither* . . . *nor* . . ., etc.). Thus *at least two and not more than ten* is intersective (in fact cardinal) because *at least two* and *more than ten* are. *Either all but two or all but three* is co-intersective (in fact co-cardinal) because each disjunct is, and *less than one out of ten or else more than nine out of ten* is proportional, since again each disjunct is.

Prop 6 Pointwise Boolean functions of proportional, (co)-intersective, and (co-) cardinal functions are proportional, (co)-intersective, and (co-)cardinal respectively.

We now return to sortal reducibility. Theorem 7 (Keenan 1993) characterizes the sortally reducible Dets (assumed conservative). See the appendix for a proof sketch.

Theorem 7 A Det_1 function D is sortally reducible iff D is intersective or D is co-intersective. Thus the classes of Dets that generalize the standard existential and universal quantifiers are *just* those that satisfy sortal reducibility. All other Dets make essential use of the noun restriction on the domain of quantification.

2.3. Generalizing Dets and DPs

2.3.1. k-Place Dets

Following Keenan (1987) and Keenan and Westerstähl (1997), we treat the underlined expressions in (24) as Det_2 s: They combine with two Ns to form a DP and semantically map pairs of sets to GQs.

- (24) a. More students than teachers came to the party.
 b. Fewer students than teachers came to the party.
 c. Exactly/almost as many students as teachers . . .
 d. More than ten times as many students as teachers . . .
 e. How many more students than teachers were arrested?
 f. The same number of students as teachers . . .
 g. A larger number of students than teachers . . .
 h. An equal number of students and teachers . . .

(25) gives the denotation of the Det_2 *more . . . than* . . ., with the N arguments in the easy-to-read order. Most of the other denotations in (24) are

defined similarly. Read (25) as “More A_1 s than A_2 s have B iff the number of A_1 s with B is greater than the number of A_2 s with B.”

$$(25) (\text{MORE } A_1 \text{ THAN } A_2)(B) = \text{True iff } |A_1 \cap B| > |A_2 \cap B|$$

Generalizing, a Det_k combines with a k-tuple of Ns to form a DP and semantically maps k-tuples of sets to GQs. Does English present Det_k s for $k > 2$? Perhaps. It is reasonable, for example, to treat *every...and...* as a k-place Det, all $k > 0$. (26) interprets it as a Det_3 .

- (26) a. every man, woman, and child jumped overboard
 b. $(\text{EVERY} \dots \text{AND} \dots)(A, B, C)(D) = \text{EVERY}(A \cup B \cup C)(D)$

The *and* in (26a) cannot be taken to merely coordinate the Ns, otherwise (26a) would mean that every object that was simultaneously a man, woman, and child jumped overboard, which is empirically incorrect. The pattern whereby the value of a Det_k built from a Det_1 and *and* maps the n noun properties to whatever the Det_1 maps their union to, is fairly general:

Def 7 For $D \in [P_E \rightarrow GQ_E]$, $(D \dots \text{AND} \dots)$ maps $(P_E)^k$ into GQ_E , all k , defined by:

$$(D \dots \text{AND} \dots)(A_1, \dots, A_k) = D(\cup_{1 \leq i \leq k} A_i)$$

- (27) a. About fifty men and women jumped overboard
 $(\text{ABOUT FIFTY} \dots \text{AND} \dots)(A, B)(D) = (\text{ABOUT FIFTY})(A \cup B)(D)$
 b. The fifty boys and girls laughed at the joke
 $(\text{THE FIFTY} \dots \text{AND} \dots)(A, B)(D) = (\text{THE FIFTY})(A \cup B)(D)$
 c. Which boys and girls came to the party?
 $(\text{WHICH} \dots \text{AND} \dots)(A)(B)(D) = \text{WHICH}(A \cup B)(D)$

Keenan and Moss (1985) offer other examples. The most convincing case that we need Det_2 s in the representation of English comes from the examples (24) as they do not lend themselves to a syntactic or semantic reduction to compounds of Dets of lesser arity.

The Det_2 s in (24) are all cardinal, meaning that the value of the DP they build at a predicate property B is decided by the cardinality of the intersection of B with each noun property. Defining conservativity, (co-)intersectivity, and (co-)cardinality for k-place Dets involves no real change from the unary case: A cardinal Det_k decides truth by checking the cardinality of the intersection of the predicate property with each of the k Noun properties. An intersective Det_k checks the intersection of that property with each N property, and so on:

Def 8 For all k , all functions D from k-tuples of properties over E into $[P_E \rightarrow X]$,

- a. D is *cardinal* iff for all k-tuples A, A' of properties and all properties B, B' if $|A_i \cap B| = |A'_i \cap B'|$, all $1 \leq i \leq k$, then $DAB = DA'B'$
 b. D is *conservative* iff for all k-tuples $A = \langle A_1, \dots, A_k \rangle$ and all properties B, B' , if $A_i \cap B = A_i \cap B'$, all $1 \leq i \leq k$, then $DAB = DAB'$

Define *intersectivity* for Det_k s by eliminating from (a) the cardinality signs; *co-cardinal* and *co-intersective* for Det_k s by replacing $A_i \cap B$ with $A_i - B$. For some additional types of Dets, see Keenan and Westerståhl (1997). We now generalize our notion of generalized quantifier.

2.3.2. Crossing the Frege Boundary: DPs as Arity Reducers $P_{n+1} \rightarrow P_n$

We have presented DPs syntactically as subjects, combining with P_1 s to form P_0 s (Ss). But of course they also combine with P_2 s to form P_1 s and in general with P_{n+1} s to form P_n s.

- (28) a. John interviewed every candidate
 b. He interviewed more men than women
 c. He interviewed every candidate but Mary
 d. We told every detective several lies
 e. We sent no offender more than two warnings

In (28a,b,c) we treat *interview* as a binary relation denoting P_2 . And *interviewed every candidate* is a P_1 whose denotation is the set of objects that stand in the INTERVIEW relation to each candidate. To obtain these interpretations we now treat GQs as functions F whose domain is the set of $n+1$ -ary relations over E and whose value at any such R is an n -ary relation over E . Writing R_n for $P(E^n)$, the set of n -ary relations over E , we define:

Def 9 F is an *arity-1 reducer* over E iff F is a function with domain $\cup_n R^{n+1}$ satisfying, for all n -tuples a ,

$$A \varepsilon F(R) \text{ iff } F(aR) = \text{True}, \quad \text{where } aR =_{\text{def}} \{b | (a,b) \varepsilon R\}$$

Note: The value of F at an $n+1$ -ary relation is an n -ary relation uniquely determined by the value of F at sets (unary relations). Thus we compute the interpretation of (29) directly:

$$\begin{aligned} (29) \quad & \text{No teacher} \quad \text{interviewed} \quad \text{every student} \\ & \text{NO(TE)} \quad \text{INTERVIEW} \quad \text{EVERY(ST)} \\ & \quad \text{EVERY(ST)(INTERVIEW)} \\ & \quad \text{NO(TE)(EVERY(ST)(INTERVIEW))} \\ & = \text{NO(TE)}(\{a | \text{ST} \subseteq a\text{INTERVIEW}\}) \\ & = \text{True iff } \text{TE} \cap \{a | \text{ST} \subseteq a\text{INTERVIEW}\} = \emptyset \end{aligned}$$

Thus (29) is true iff TEACHER has an empty intersection with the set of individuals whose interviewees include all the students.

Thus the arity-1 reducers are just the extensions of GQs to $n+1$ -ary relations generally. And we now interpret DPs as arity-1 reducers. This interpretation of DPs crosses a certain Frege boundary in an apparently innocuous way, but one that opens up many new avenues of generalization. Observe first that the standard Fregean solution to the problem of multiple quantification consists of iterated application of unary

$$(34) \text{ SELF}(R) = \{a \in E \mid (a, a) \in R\}$$

Probably SELF is a “new” function, not an extension of a GQ that maps unary relations to truth-values. Comparable claims hold for infinitely many other complex anaphors: *every student but himself*, *only himself* and *the teacher*, and so on.

2.4. “(Non)-Logical” Dets

Our classes of Dets above have quietly included a few, usually complex, expressions, such as *more male than female*, *John's ten*, *no . . . but John*, and so on, which lack the “logical” character of classical quantifiers. This character is (partially) captured by the idea that “logical” expressions are insensitive to which individuals constitute the domain E . The proper formulation of this idea is global, involving comparison of models with different domains, but the global idea also manifests itself locally. Here first is a local logicity condition that Det_1 denotations may meet. Then we see that the notion applies to expressions of all categories.

Def 10 For π be a permutation of E (a bijection from E to E), π extends to a permutation of P_E by setting $\pi(A) = \{\pi(x) \mid x \in A\}$. Then a Det_1 function D is *permutation invariant* (PI) iff for all such bijections π , $D(\pi A)(\pi B) = DAB$, all $A, B \subseteq E$.

Fact: Set theoretic and cardinality properties are preserved by such permutations π . For example:

$$(35) \text{ For all subsets } A, B \text{ of } E, A \subseteq B \text{ iff } \pi A \subseteq \pi B, \text{ and } |A| \leq |B| \text{ iff } |\pi A| \leq |\pi B|$$

So all the functions in (3) are PI. For example, $\text{ALL}(A)(B) = \text{True}$ iff $A \subseteq B$, iff $\pi A \subseteq \pi B$, iff $\text{ALL}(\pi A)(\pi B) = \text{True}$, so $\text{ALL}(A)(B) = \text{ALL}(\pi A)(\pi B)$. Similarly for the other Dets whose denotations are defined solely in terms of set theoretical and cardinality notions.

But Dets like *more male than female*, *John's ten*, *no . . . but John*, and so on may denote many different functions for a given E . Usually these functions are not PI (though PI ones are not ruled out). For example, for each way of interpreting *male* and *female* we have a different way of interpreting *more male than female*. *no . . . but John*, defined in (36), has as many denotations as there are elements in the domain. So in an E with at least two individuals, say John and Sam, *no . . . but John* will not be PI: Let π interchange John and Sam and map everything else to itself. Then (36b) is True when $A = \{\text{John}\}$ and

- (36) a. No student but John came to the lecture
 b. $(\text{NO } A \text{ BUT JOHN})(B) = \text{True}$ iff $A \cap B = \{\text{John}\}$

$B = \{\text{John}, \text{Sam}\}$. But $\pi(A) = \{\text{Sam}\}$ and $\pi(B) = \{\text{Sam}, \text{John}\}$, so $\pi A \cap \pi B = \{\text{Sam}\}$, so $(\text{NO } \pi A \text{ BUT JOHN})(\pi B) = \text{False}$, whence $\text{NO } \dots \text{BUT JOHN}$ is not AI.

Thus, contra Etchemendy (1990), we have a property of denotations that distinguishes among logical and nonlogical Dets, a property not in fact peculiar to Dets as the invariance conditions above extend directly to expressions of all categories. The basic idea is that “logical” expressions are ones whose denotations are unchanged (invariant) under the structure preserving changes (automorphisms) of the semantic primitives. In the minimalist extensional semantics we are using here the only semantic primitives of a model structure are a set E with no assumed structure and the set $\{\text{True}, \text{False}\}$ with its Boolean structure (given, for example, by the standard truth functions). In more extensive models for English we would surely have, for example, a set of time points (or intervals) with an ordering relation, a set of spatial points (or regions) with a distance relation, and so on. Now, an automorphism of a structured set $(A, \dots)$ is simply a bijection from A to A , which does not change the structure. For example, given a relational structure (A, R) an automorphism π of (A, R) is a permutation of A that fixes R (xRy iff $\pi x R \pi y$, which we more succinctly write simply as $\pi R = R$). If an initial set, like E , comes with no structure, an automorphism is simply a permutation of that set. But of the two permutations of the set $\{\text{True}, \text{False}\}$ of truth-values, only the identity map is an automorphism, since only that permutation fixes the truth tables for *and* and *or*.

Now given a collection C of (structured) sets, assumed disjoint for simplicity, a *basic automorphism* π of these sets is a function that is an automorphism of each of these sets (when restricted to that set). π extends to higher types as follows:

- (37) a. for R an n -ary relation, $\pi R =_{\text{def}} \{\pi x | x \in R\}$, where $\pi(<x_1, \dots, x_n>) =_{\text{def}} <\pi(x_1), \dots, \pi(x_n)>$.
 b. for F a map from X to Y , πF maps each πx to $\pi(F(x))$.
 (38) An element G in the type hierarchy generated by a collection of structured sets is *automorphism invariant* (AI) iff $\pi(G) = G$, all basic automorphisms π .

Iterating (37b) says that a Det_1 denotation D is AI iff for all $A, B \subseteq E$, all basic automorphisms π , $D(\pi A)(\pi B) = \pi(DAB)$, and $\pi(DAB) = DAB$, since DAB is a truth-value.

Expressions we call *logical constants* always denote AI objects, and in fact satisfy a stronger, global condition (see section 3). Even so, Alness provides a denotational property of *constants*, again contra claims made in the literature (Etchemendy 1990). To see this let us formalize the notion of a (minimal, extensional) model M standardly as a triple $(E, 2, m)$, where E is a nonempty set, 2 is the two-element set $\{\text{True}, \text{False}\}$ with its Boolean structure, and m is a function that interprets nonderived expressions as appropriate elements of the type hierarchy built from E and 2 . For example, if d is a P_1 then $m(d)$ is a subset of E , and so on. Then we define an expression to be a *constant* if it always has the same denotation

in models with the same domain (Keenan and Stavi, 1986). And we define an expression to be (locally) *logical* if it always denotes an AI element of its type:

Def 11 a. An expression d is a *constant* iff for all models $(E, 2, m)$ and $(E, 2, m')$, $m(d) = m'(d)$.

10 b. An expression d is *logical* iff for all models $(E, 2, m)$, $m(d)$ is AI

Prop 8 Constants always denote AI elements of their type.

Proof Let d be a constant. We must show for all models $(E, 2, m)$ that $m(d)$ is AI. Let π be an arbitrary permutation of E . Then $m(d) = (\pi \circ m)(d)$, d is constant, $= \pi(m(d))$, def. of composition. Thus $m(d) = \pi(m(d))$ all π , so $m(d)$ is AI⁵.

Prop 8 shows that logicity and constancy are not independent. But equally they are not equivalent, as logicity does not entail constant interpretation. An expression might admit of many denotations over a given domain, all of which were AI. It would then be logical as defined but not constant. A candidate class of such expressions is given in (39).

(39) about/approximately a hundred, almost/nearly a hundred, all but about ten, almost nine out of ten

One way to account for the vagueness of these Dets would be to allow models $(E, 2, m)$ and $(E, 2, m')$, where m maps *about a hundred* to BETWEEN 98 AND 102, m' maps it to BETWEEN 97 AND 103, and so on. So each function denoted by *about a hundred* is AI, but the set of possibilities for a domain with at least 103 elements is more than a unit set. Thus *about a hundred* is logical—it always denotes an AI function, but it is not constant—it may have different denotations in the same domain. Of course in CL the standard quantifiers are constants (and thus logical). So natural languages differ from CL in presenting Dets that are logical but not constant, like *about a hundred*, as well as ones that are nonlogical (hence nonconstant), such as *more male than female*, *no... but John*, and *Which*?

These remarks are helpful in teasing apart the properties that constitute expressions traditionally called *logical constants*. But the traditional notion is richer than merely being both logical and constant. Consider for example a Det₁ expression *blik* whose interpretation in any model $(E, 2, m)$ is given in (40). where $\mathbf{0}$ is that Det₁ function mapping all A, B to False:

5. Note that the proof does assume that the class of models is closed under auto-morphisms, that is, that $\pi \circ m$ is an acceptable interpretation, any permutation π of E . This condition is always met, and it would be bizarre indeed to rule out some isomorphic images of models from the class of models, especially given the general theorem that isomorphic models make the same Ss true (= are elementarily equivalent); that is, our language is not rich enough to discriminate among isomorphic models.

$$(40) m(\text{blik}) = \begin{cases} \text{ALL if } 7 \in E \\ 0 \text{ otherwise} \end{cases}$$

In each model *blik* has just one denotation, so it is constant, and that denotation is always AI, so *blik* is logical. But *blik* is not counted as logical constant, since it can tell whether a particular entity is in the domain or not. To improve our characterization of logical constants we move to a global perspective. That perspective also enables a better understanding of the role of the Noun property in restricting the domain of quantification, and is helpful in studying logical expressive power—the sense in which English Dets enable us to distinguish among models indistinguishable in CL.

3. GLOBAL PROPERTIES OF ENGLISH QUANTIFIERS

3.1. Global Logicality

The definition of automorphism invariance in (37) and (38) extends in an obvious way to *isomorphism invariance* simply by letting π be a *basic isomorphism*, that is, π restricted to each of the initial structured sets is an isomorphism (whose range is disjoint from that of π restricted to any of the other initial structured sets). Limiting ourselves to models of the sort considered so far, we have:

Def 12 An expression d is *isomorphism invariant* (ISOM) iff for all models $(E, 2, m)$ and $(E', 2, m')$ and all bijections $\pi: E \rightarrow E'$, $m'(d) = \pi(m(d))$

Prop 9 If d is ISOM then d is constant and $m(d)$ is AI, all models $(E, 2, m)$.

Thesis Logical constants are syntactically simple expressions that are ISOM

The Thesis guarantees that logical constants are both logical and constant, and rules out cases like *blik* above, since in $E = \{a, b\}$ $m(\text{blik}) = \text{ALL}$ but in the isomorphic $\{7, 8\}$ $m'(\text{blik}) = 0$ and no bijection π from $\{a, b\}$ to $\{7, 8\}$ can map ALL to 0. The Thesis is not without problems, however. It rules out *about a hundred* as interpreted earlier, since it is not constant.⁶ More problematically, it lets in interpretations that vary from domain to domain when the conditioning factor is itself ISOM. For example, it allows us to interpret an expression as ALL if the domain is finite, and as INFINITELY MANY if the domain is infinite. Finally, as with our earlier characterization, this one effectively identifies *logical constant* with *set theoretical constant*, which most logicians would probably consider too generous. But we do not pursue this issue further here.

6. We might avoid this problem by generalizing to allow expressions to determine sets of possible denotations and require the set denoted in an isomorphic image be the set of images of the elements of the set (see Keenan and Westerståhl 1997). Alternatively we might try to account for the vagueness in *about a hundred* differently.

3.2. The Noun Property Revisited

Conservativity just guarantees (in the simplest case) that in evaluating DAB we can ignore elements of B that do not lie in A. But it does not guarantee that we can ignore all non-As. Thus BLIK below is conservative, and ISOM:

$$(41) \text{ BLIK}(A)(B) = \text{True iff } |E - A| = 3$$

Blik is unnatural, since we could change the truth-value of, for example, *Blik cats can fly* merely by adding some non-cats to the domain. Van Benthem (1984) was the first to formulate the domain independence condition (called *Extension* by him), which natural language Dets seem to satisfy. Namely, a Det is interpreted as a functional D which associates with each domain E a function D_E satisfying (42).

$$(42) \text{ Domain Independence (DI): For all domains } E, E' \text{ with } A, B \subseteq E, E', \\ D_E AB = D_{E'} AB$$

BLIK fails DI since $\text{BLIK}\emptyset\emptyset = \text{True}$ for $E = \{a, b, c\}$ but False for $E' = \{a, b, c, d\}$. Note that DI and CONS are independent: BLIK above is CONS but not DI; the Härtig quantifier ($\text{HAB} = \text{True iff } |A| = |B|$) is DI but fails CONS in each E (with $|E| > 1$). It seems then that

$$(43) \text{ Natural language Dets are both domain independent and conservative.}$$

Keenan and Westerstähl (1997) state a different condition in terms of *relativization* that captures domain independence and conservativity as a single condition (for Det_k s).

The notion of domain independence extends naturally to expressions of most if not all categories. Det_k s for $k > 1$ are notably DI. But in distinction to ISOM it is not a property that holds of all operators in CL. Notably $\forall$ defined earlier is not DI. Nor is the English DP *all entities*. So as a general constraint DI is perhaps limited to Det_k s, $k \geq 1$.

Finally, the notion of DI helps us to better understand the relation between the cardinal and merely intersective Dets, as well as that between the co-cardinal and co-intersective ones:

Theorem 10 In the presence of DI, for each E,

- a. D is cardinal iff D is intersective and D is automorphism invariant
- b. D is co-cardinal iff D is co-intersective and automorphism invariant

3.3. Expressive Power: First-Order Definability

Because first-order logic has several logically appealing properties, notably the syntactic characterizability of entailment and validity, there has been considerable interest in studying whether a proper semantics for English could be given in first order. Various scholars (Hintikka 1974; Gabbay and Moravcsik 1974; Boolos 1984) have argued in the negative, using often subtle facts from English. Here, generalizing from early work

by Kaplan (1966) and later work by Barwise and Cooper (1981), we see that the range of overtly expressible quantifier expressions goes well beyond those definable in first order. Let us first consider English Dets that are first-order definable.

First-order formulas are those built from n -place predicate symbols followed by n variables, closed under the formation of Boolean compounds (*and*, *or*, *not* will do) and standard quantification: $\forall x\Phi$ and $\exists x\Phi$ are in the language for each variable x and each first-order formula Φ . A *first-order sentence* is a first-order formula with no free occurrences of variables (that is, every occurrence of a variable x is either in a quantifier $\forall x$ or $\exists x$, or it is in a subformula Φ' of Φ to which a quantifier $\forall x$ or $\exists x$ has been prefixed).

Let D associate with each domain E a map D_E from P_E into $\{P_E \rightarrow \{\text{True}, \text{False}\}\}$. Then D is *first-order definable* (fod) iff there is a first-order sentence Φ whose nonlogical constants are just two one-place predicate symbols P, Q such that Φ is true in a model $M = (E, 2, m)$ iff $D_E(mP)(mQ) = \text{True}$. First-order definability of Det_k s is defined analogously. Note that a fod Det_k is isomorphism invariant, since the defining sentence has the same truth-value in isomorphic models.

For example, the intersective AT LEAST TWO is definable by $\Phi = \exists x\exists y(x \neq y \ \& \ Px \ \& \ Py \ \& \ Qx \ \& \ Qy)$. AT LEAST n is defined analogously. MORE THAN n is just AT LEAST $n + 1$. EXACTLY n is $(\text{AT LEAST } n \wedge \neg \text{MORE THAN } n)$, AT MOST n is $\neg \text{MORE THAN } n$, and so on. Thus

Claim 2 The finite cardinal Det_1 s are first-order definable; so are the finite co-cardinal Det_1 s.

The finite co-cardinal Det_1 s are *all but n* and *all but at most n* , for n a finite cardinal. Note that *All but two As are Bs* is paraphrased by *Exactly two As are not Bs*, which is fod since *Exactly two* is.

We see then that the “finite” intersective and co-intersective Det_1 s are fod. As is well known, however, (co)-intersective Det_1 s like *infinitely many*, *just finitely many*, *all but finitely many* are not fod. The claim that they take English outside of first order is weak, however, as they rely on a technical notion of (in)finite that cannot be assumed to be part of a normal speaker’s competence. Still, to get a feeling for what is fod, it is instructive to consider the proof that these Dets are not fod. The tack taken is to find a property that first-order languages have that is lost if any of these quantifiers is added to the language. Hence that quantifier cannot be defined within first order.

The property that is chosen for these cases is *compactness*: For S a set of sentences (in a first-order language), there is a model in which all the sentences in S are True iff for every finite subset K of S , there is a model in which all the sentences in K are True. Succinctly: S has a model iff every finite subset of S has a model. Now we show that *all but finitely many* is not fod by showing that the language that results by enriching a first-order language with this quantifier fails to be compact. Let L be first

order with P, Q one-place predicate symbols and add in the quantifier *for all but finitely many* x . Let K' be the infinite set of sentences $\{\text{There are at least } n \text{ Ps which are not Qs} \mid 1 < n \in \mathbb{N}\}$. (We have already seen how to say "There are at least n Ps which are not Qs" above.) And set $K = K' \cup \{\text{for all but finitely many } x (Px \ \& \ \neg Qx)\}$. Clearly each finite subset H of K has a model: Choose a model M that interprets Q as the empty set and P with n elements, n the largest number such that the sentence *There are at least n Ps which are not Qs* is in H . (If no such sentence is in H interpret P as empty as well.) Clearly M is a model of H , whether *for all but finitely many* $x (Px \ \& \ \neg Qx)$ is in H or not. So every finite subset of K has a model. But clearly K itself does not, since for any n there are more than n Ps which lack Q , hence the number of Ps which lack Q cannot be finite. So adding *all but finitely many* takes us outside of first order.

We turn now to the cases of interest to us: the proportionality quantifiers and then the k -place (co)-cardinal Dets. In both cases their typical members are not fod. Kaplan (1966) first pointed out that *most* in the sense of *more than half* was not fod. But the argument used infinite sets in an essential way. Later Barwise and Cooper (1981) showed that *most* was not fod even in finite models. Here the proof strategy cuts deeper than the compactness argument above: For each first-order Φ whose nonlogical symbols are just two P_1 symbols they show how to construct two models that make the same first-order sentences true, but *Most Ps are Qs* is True in one and False in the other, so *most* cannot be first-order definable.

Barwise and Cooper treat MOST as we did in (3): It maps A, B to True iff $|A \cap B| > (1/2) \times |A|$. Their proof can be modified to hold when $1/2$ is replaced by n/m , $0 < n < m$, and when $>$ is replaced by $<$. Thus *at least/exactly/more than two out of three*, and so on are not fod, nor, obviously are their complements *less than/at most/not more than seven out of ten*, and so on.

Let us refer to these Dets as *properly proportional*. Then,

Claim 3 The properly proportional Dets are not fod, and

Claim 4 The nontrivial cardinal Det₂s are not fod

To see that *more...than...* is not fod observe that $(\text{MORE } (A \cap B) \text{ THAN } (A - B))(E) = \text{MOST}(A)(B)$, so if *more...than...* were fod *most* would be also, and it is not.

In sum, the case that semantic interpretation of English cannot be given in first order is quite solid.⁷ Basic proportional expressions, such as fractions and percentage ones—*just three out of ten*, *not more than one...in ten*, *less than ten per cent of the*, and so on. are not fod. Nor are basic cardinal comparatives: *I know more men than women / just as many men as women*, and so on.

7. Since Barwise and Cooper (1981) techniques for proving undefinability results in languages with generalized quantifiers have developed, leading to stronger results. See Westerståhl (1989, 1991) and Kolaitis and Väänänen (1995). The latter show that the properly proportional Dets are not definable in terms of any logic with any finite number of monotone unary quantifiers (result cited from Keenan and Westerståhl (1997)).

APPENDIX

1. Proof of Proposition 1

$\Rightarrow$ Let D be conservative. Then since $A \cap B = A \cap (A \cap B)$, $DAB = DAA \cap B$.

$\Leftarrow$ Let $A \cap B = A \cap B'$. Show $DAB = DAB'$. But $DAB = DAA \cap B = DAA \cap B' = DAB'$.

2. Proof of Propositions 3a and 3b.

Prop 3a Let D be intersective. Then since $A \cap B = (A \cap B) \cap E$, then $D(A)(B) = D(A \cap B)(E)$. Now let $D(X)(Y) = D(X \cap Y)(E)$, all $X, Y \subseteq E$. Let $A \cap B = A' \cap B'$. Then $DAB = D(A \cap B)(E) = D(A' \cap B')(E)$, by the assumption, $= DA'B'$, completing the proof.

Prop 3b Let D be intersective. Then since $A \cap B = E \cap (A \cap B)$ then $DAB = D(E)(A \cap B)$. Going the other way, assume the right-hand side of (2b). Let $A \cap B = A' \cap B'$. Then $DAB = D(E)(A \cap B) = D(E)(A' \cap B') = DA'B'$, completing the proof.

3. Proof of Theorem 6. We use the precise definition of *sortally reducible*:

Def 1 D is *sortally reducible* iff there is a two-place Boolean function h such that for all properties A, B

$$DAB = D(E)(h(A, B))$$

The sixteen Boolean functions of two places are given in set theoretic form by setting $h(A, B)$ to be any of the sixteen options listed below (the second eight are just the Boolean complements of the first):

1. E 2. A 3. B 4. $A \cdot B$ 5. $A \cdot \neg B$ 6. $\neg A \cdot B$ 7. $\neg A \cdot \neg B$ 8. $(A \cap B) \cup (\neg A \cap \neg B)$
9. $\emptyset$ 10. $\neg A$ 11. $\neg B$ 12. $\neg A \cup \neg B$ 13. $\neg A \cup B$ 14. $A \cup \neg B$
15. $A \cup B$ 16. $(\neg A \cup \neg B) \cap (A \cup B)$

So for each n between 1 and 16, h_n is that function mapping A, B to the entry next to n above. The proof of the theorem now proceeds by cases. We do a few representative ones and refer the reader to Keenan (2000) for the others:

1. Suppose first that D reduces via h_1 (that function mapping all A, B to E). Then $DAB = D(E)(h_1(A, B))$, by the assumption, $= D(E)(E)$ by the definition of h_1 , so D is constant, either mapping all A, B to False or all A, B to True. In either case it is intersective (and co-intersective, since it is trivial). Similarly, one checks that if D reduces via h_9 it is trivial.
2. Suppose that D reduces via h_2 . Then $DAB = D(E)(h_2(A, B))$ by the assumption, $= D(E)(A)$, by the definition of h_2 , $= D(E)(h_2(E, A))$, by the assumption, $= DEE$, by the definition of h_2 . So again D is constant and thus intersective (and co-intersective). Similarly, if D reduces via h_{10} then $DAB = DE\emptyset$ so again D is constant.
3. Let D reduce via h_3 . Then $DAB = D(A)(A \cap B)$, since D is conservative, $= D(E)(h_3(A, A \cap B))$, by the assumption, $= D(E)(A \cap B)$, by the def of h_3 , whence D is intersective (if $A \cap B = A' \cap B'$ then $DAB = D(E)(A \cap B) = D$

$(E)(A' \cap B') = DA'B')$. Equally if D reduces via h_{11} it is intersective (since if $A \cap B = A' \cap B'$ then $\neg A \cup \neg B = \neg(A \cap B) = \neg(A' \cap B') = \neg A' \cup \neg B'$).

Similarly, if D reduces via h_4 or h_{12} then D is intersective; if it reduces via h_8 or its complement h_{16} then it is co-intersective, and if it reduces via h_6 , h_7 , or their complements h_{14} or h_{15} it is constant.

REFERENCES

- Barwise, J., and R. Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4:159–219.
- Boolos, G. 1984. Non-first orderizability again. *Linguistic Inquiry* 12:465–467.
- Etchemendy, J. 1990. *The Concept of Logical Consequence*. Cambridge, Mass.: Harvard University Press.
- Frege, G. 1892. On concept and object. In *Translations from the Writings of Gottlob Frege*, ed. P. Geach and M. Black, 42–56. Oxford: Blackwell, 1952.
- Gabbay, D., and J. Moravcsik. 1974. Branching quantifiers and Montague grammar. *Theoretical Linguistics* 1:139–157.
- Groenendijk, J., and M. Stokhof. 1997. Questions. In *The Handbook of Logic and Linguistics*, ed. J. van Benthem and A. ter Meulen, 1055–1125. Amsterdam: North Holland.
- Hintikka, J. 1973. Quantifiers vs. quantification theory. *Dialectica* 27:329–358. Reprinted in *Linguistic Inquiry* 5:153–177.
- Kaplan, D. 1966. “Rescher’s plurality-quantification” and generalized plurality quantification. *Journal of Symbolic Logic* 321:153–154.
- Keenan, E. L. 1987. A semantic definition of “Indefinite NP.” In *The Representation of (In)Definiteness*, ed. E. Reuland and A. ter Meulen, 286–317. Cambridge, Mass.: MIT Press.
- Keenan, E. L. 1989. Semantic case theory. In *Semantics and Contextual Expression*, ed. R. Bartsch, J. van Benthem, and P. van Emde Boas, 33–57. Dordrecht: Foris.
- Keenan, E. L. 1992. Beyond the Frege Boundary. *Linguistics and Philosophy* 15:199–221.
- Keenan, E. L. 1993. Natural language, sortal reducibility and generalized quantifiers. *Journal of Symbolic Logic* 58(1):314–325.
- Keenan, E. L. 2000. Quantification in English is inherently sortal. *History of Philosophy and Logic* 20: 251–265.
- Keenan, E. L., and L. S. Moss. 1985. Generalized quantifiers and the expressive power of natural language. In *Generalized Quantifiers in Natural Language*, ed. J. van Benthem and A. ter Meulen, 73–127. Dordrecht: Foris.
- Keenan, E. L., and J. Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9:253–326.
- Keenan, E. L., and D. Westerståhl. 1997. Generalized quantifiers in linguistics and logic. In *The Handbook of Language and Logic*, ed. J. van Benthem and A. ter Meulen, 837–893. Amsterdam: Elsevier.
- Kolaitis, Ph., and J. Väänänen. 1995. Generalized quantifiers and pebble games on finite structures. *Annals of Pure and Applied Logic* 74:23–75.
- Russell, B. 1908/1956. Mathematical logic as based on the theory of types. In *Logic and Knowledge Essays 1901–1950*, ed. R. C. Marsh, 57–103. London: George Allen and Unwin.

- van Benthem, J. 1984. Questions about quantifiers. *Journal of Symbolic Logic* 49:443–466.
- Westerståhl, D. 1989. Quantifiers in formal and natural languages. In *The Handbook of Philosophical Logic*, ed. D. Gabbay and F. Guentner, 4:1–131. Dordrecht: Reidel.
- Westerståhl, D. 1991. Relativization of quantifiers in finite models. In *Generalized Quantifier Theory and Applications*, ed. J. van der Does and J. van Eick, 187–205. Amsterdam: Institute of Logic, Language and Information.

CHAPTER 6

The Lesson of Kaplan's Paradox about Possible World Semantics

C. Anthony Anderson

1. THE PROBLEM

David Kaplan has shown that the natural extension of the now-standard possible world semantics for modal propositional logic invalidates the principle:

$$(A) \quad \forall p \Diamond \forall s (Qs \leftrightarrow p = s).^1$$

The formula itself would belong to what could be called "extended modal-propositional calculus with identity; first-order and singulary with-respect-to-propositions, applied or pure." To modal propositional logic one adds quantifiers with propositional variables, identity as standing between propositional variables (and perhaps complex well-formed formulas) and at least one constant or variable '*Q*' having the syntax of a singulary sentential operator. '*Q*' may with equal justice be thought of as conveying some property of propositions. I will assume, to simplify the discussion, that the underlying modal logic is S5. In that context the villain of the piece may be taken to be:

$$(A^*) \quad \forall p \Diamond \forall s [Qs \leftrightarrow \Box(p \leftrightarrow s)]$$

as Kaplan notes in passing. One danger in this simplifying assumption is that someone might think that S5 is somehow the source of the trouble. But it is not. Furthermore, (A^*) has less grip on one's logical intuitions than (A) . Still, since possible world semantics identifies necessarily equivalent propositions, it is certain that (A^*) is to be seen as following from (A) . And since (A^*) is already a well-formed formula of a simpler extension of propositional modal logic, in that context we do not need to

1. There is a typographical error in the statement of this principle on page 43 of the article "A Problem in Possible World Semantics," in *Modality, Morality and Belief*, ed. Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher (New York: Cambridge University Press, 1995). I have made trivial changes of bound variables.

take identity of propositions as primitive. Furthermore, the problem seen as concerning (A^*) takes on a striking and enlightening aspect in that context.

Given the intended possible world semantics, there is a *translation* of (A^*) into a classical setting—ordinary extensional higher-order logic, third-order logic, to be specific, but with a single third-order predicate. The idea is itself due to David Kaplan.² The translation works like this. A set of possible worlds is just a nonempty set. The propositions are subsets of it—elements of its power set. Necessity just means “true” at every element of the given set. Without having to detour through the modal semantics, we can just treat propositional variables as one-place predicate variables and necessity as “truth” at every element in the individual domain. The singular predicate ‘ Q ’ becomes a binary (third-order) predicate ‘ Q ’, say. Then the translation of (A^*) is:

$$\forall P \exists x \forall S [Q(S, x) \leftrightarrow \forall x (Px \leftrightarrow Sx)]$$

(The direct translation of our original (A) would be: $\forall P \exists x \forall S [Q(S, x) \leftrightarrow (P = S)]$ Now by a direct application of Cantor’s argument, as it would be expressed in this logic, this formula can be refuted—that is, its negation can be shown to be a theorem in a standard formulation of third-order logic. The intuitive idea behind it is that the relation Q would provide a one-to-one correspondence between the sets of elements in the individual domain and some of the elements of that domain, contrary to Cantor’s Theorem.

This just emphasizes Kaplan’s observation that we have here a *cardinality* problem. The reasoning that Kaplan uses in the metalanguage, itself a cardinality argument, is here just rendered directly as an argument in higher-order logic.³

The upshot of this exercise is just some further confirmation that the invalidity of the Kaplan formula is a firm fact, built directly into the semantics by way of Cantorian cardinality constraints.

2. METAPHYSICAL MEDDLING IN LOGIC

Kaplan argues with considerable force that metaphysics should not meddle with logic. However lovely the resulting picture of possible world metaphysics, he says that it should not be allowed to constrain, by way of possible world semantics, our modal logic to the point of forcing upon us such things as the invalidity of (A) . So, he urges us to conclude, the

2. “S5 with Quantifiable Propositional Variables,” *Journal of Symbolic Logic* 35 (1970): 355.

3. Translating in the other direction is also entertaining. For example, the second-order theorem $\exists! P \forall x P(x)$ yields $\exists! p \Box p$, a valid formula in standard possible world semantics. Familiarity with this last formula has apparently assuaged the contempt it so richly deserves.

paradox should not be resolved by just accepting the logical invalidity of (A).

Here Kaplan seems to be fighting for a lost cause. The sentiment that logic should be the neutral arbiter of philosophical disputes is a noble one, and everyone should firmly endorse its spirit. But given the bizarre multitude of metaphysical principles that have been endorsed by some philosopher or other, the thing is hopeless when uniformly applied in the flesh. Leon Trotsky and Alfred Korzybski both argued against ' $A = A$ '.⁴ And Graham Priest has argued that some things can be both true and false. Even logic cannot be neutral about all these things. And if the logic is firm enough, the metaphysics may have to be pronounced crack-brained. If the metaphysics is firm enough, the logic may have to go.

It is difficult not to conclude that David Lewis was right when he said in response to Kaplan, "[T]he way of the neutralist is hard."⁵

3. THE EPIMENIDES AND THE THEORY OF PROPOSITIONS

The dialectic takes what should be seen as a new twist when an additional argument is offered that seems to cause further problems for (A). Consider the proposition expressed by $\forall s (Qs \rightarrow \sim s)$. Given (A), a contradiction results. This is really a modified version of the classical Epimenides paradox—not quite the same as the Liar but in the same category.⁶ This is presented in a context where Kaplan is arguing the merits of ramification as a method of dealing with the difficulty posed by (A). Now this really is a difficulty for modal, and even nonmodal, logic when propositions are being directly treated. But it must be emphasized that this is something essentially different from the clash between (A) and possible world semantics. David Lewis suggested that (A) could be saved by denying that "just any set of possible worlds gives the content of some possible thought"—because the suggested choices for Q involve propositional attitudes. The idea is that propositions, taken neat as sets of possible worlds, can, should, and do invalidate (A). But the interpretation of (A) according to which it seems possible is one where propositions are seen as the contents of possible thoughts. To this Kaplan replies that this

4. Thus Alfred Korzybski, "[A] principle or premise that 'everything is identical with itself' is *invariably false to facts*"; *Science and Sanity* (Scranton, Pa.: International Non-Aristotelian Library Publishing, 1933), 194; and Leon Trotsky, "But in reality A is not equal to A"; "The ABC of Dialectical Materialism," in *Problems of Everyday Life & Other Writings on Culture and Science* (New York: Monad Press, 1973), 329.

5. *On the Plurality of Worlds*, 105n2 (Oxford: Basil Blackwell, 1986).

6. Several authors have insisted that the Epimenides, being about assertion and propositions, is essentially different from the Liar, which concerns sentences (e.g., John Myhill in "A Refutation of an Unjustified Attack on the Axiom of Reducibility," in *Bertrand Russell Memorial Volume*, ed. George W. Roberts (New York: Humanities Press, 1979)). This is correct as far as it goes, but there seem to be parallel paradoxes in virtually every case.

Epimenides paradox still shows that all is not well with possible world semantics.

But it is not clear that this new turn has anything whatever to do with possible world semantics. In any direct and natural extension of the modal propositional calculus to include quantifiers with propositional variables, identity, and constants like '*Q*', one would expect to have as a theorem:

$$(S) \exists r[r = \forall s(Qs \rightarrow \sim s)].$$

Now from this and (A) a contradiction will follow. Even from

$$(S^*) \exists r \Box[r \leftrightarrow \forall s(Qs \rightarrow \sim s)]$$

and (A^{*}) a contradiction can be deduced. The latter two are consequences of the natural and inevitable-looking axiom schema:

$$\forall p A \rightarrow S_B^p A|,$$

where *p* is a propositional variable and the wff $S_B^p A|$ is the result of substituting the wff *B* for all free occurrences of *p* in the wff *A*, unless some such occurrence is within a well-formed part of *A* of the form $\forall q C$ where *q* is a free variable of *B*. In the latter case, $S_B^p A|$ is just *A*. This is "universal instantiation" involving propositional variables and where the item substituted is a sentence or a sentential form. One would very likely include this even if a firm disbeliever in possible world semantics. The only thing that seems at all suspect is that sentences (and sentential forms) are being treated as "names" of propositions, or at least namelike. An adherent of a Fregean way of looking at the matter, according to which sentences denote truth-values and *express* propositions, may leap to the conclusion that some kind of Russellianism is the root evil here. But this is not so. In its essentials the argument goes through in a fully Frege-Churchian formulation in the logic of sense and denotation. The reasoning can be reconstructed therein by using the formalized version of "*s* is a concept of falsehood" instead of " $\sim s$." (I omit details.) The whole puzzle takes on a different complexion in that context insofar as the notions of truth and falsity appear explicitly. This suggests that the "semantical" notions, or their analogues, are the frogs at the bottom of these beer mugs.

But, again, this would be a hasty diagnosis. A close relative of the argument can be constructed in the logic of sense and denotation if it is granted that there is a "rigid" concept of negation, a concept that, to speak "possible-world-speak," picks out that concept in every possible world. (Again, I omit details.) So it looks as if the difficulty is not to be traced to Carnapian possible world semantics. It is the ghost of Epimenides. Here I would draw the conclusion that there is indeed a clash between (A) and (S), or between (A^{*}) and (S^{*}), but it does not depend on possible world semantics. Hence it seems that Kaplan's retort to Lewis

that all is not well with (A) (or (A*)) is correct, but it misses the mark insofar as he, Kaplan, sees this as a further indictment of possible world semantics.

One might observe, in opposition to Lewis, that his proposal will complicate possible world semantics by requiring, in the metalanguage, that some sets of possible worlds be singled out as those that correspond to possible thoughts. But few nowadays think that possible world semantics, as is, can handle the propositional attitudes anyway.

4. AGAINST RAMIFICATION

Kaplan concludes that the best repair for the situation is ramification. But another course has some things in its favor.

In discussions of various proposals to deal with the Liar and related paradoxes, the usual procedure seems to be that one first harshly criticizes all the alternatives to the favored one. Then when the unnatural aspects of the author's own view appear—perhaps after various maneuvers to keep them in the closet—he replies meekly that *every* proposed solution has *some* counterintuitive features. It is better to state the costs out front. In the end the whole thing is a “cost-benefit” analysis.

First note well the costs of ramification. (None of this is going to be really new.) We can no longer even express general principles about propositions. Even the hoary $\forall p \sim (p \wedge \sim p)$ goes by the board. (I have given the Russellian version.) It just is not well formed. You must specify the “order” of the propositional variable ‘*p*’. And the principles (A) and (A*), which we began by trying to save, cannot even be stated in the revised notation. We wished to allow for a property *Q* that is such that *any* proposition might have been the sole proposition having *Q*. But we are no longer allowed to speak about all propositions.

Suppose that the policeman testifies that everything the prisoner says about the case is in error and that the prisoner deposes that something the policeman says about the case is correct. Well, if the policeman is exceptionally terse in his report and the prisoner is also quite short in his remarks about the policeman, contradiction ensues. This version of the Epimenides paradox, but with a loop, is essentially due to L. Jonathan Cohen.⁷

One may complain against the ramified version of this exchange, and with some justice, that the policeman and the prisoner do not wish to confine their remarks to propositions of some order or other. Each wishes

7. In “Can the Logic of Indirect Discourse Be Formalised?” *Journal of Symbolic Logic* 22 (1957): 225–232. Some readers will recognize at once that this is very close to being a propositional version of Saul Kripke's example about Nixon and Dean (“Outline of a Theory of Truth,” *Journal of Philosophy* 72 (1975): 690–716). In fact, *that* example is about propositions when the things said are taken at face value.

to speak about everything relevant said by the other. But, according to the doctrine of ramification, they cannot. Notice that these things can be formulated in the Russell-style propositional logic with quantifiers *without any use of the notions of truth or falsehood*. We just use $\forall p(Fp \rightarrow \sim p)$ as the policeman's assertion and $\exists p(Gp \wedge p)$ for the prisoner's—where 'F' and 'G' mean "was deposited by the prisoner" and "was asserted by the policeman," respectively.

Finally, an objection urged forcefully by Fitch,⁸ Gödel,⁹ and probably many others. If the doctrine of ramification is stated as a principle about propositions, it is self-refuting. If we cannot speak about all propositions, then we cannot say that all propositions have orders. The usual response is that we just place a certain constraint on the syntax and remain mute (as we must) about the general ontological situation. But here it seems that the logician is meddling in metaphysics and will not even say why or how.

5. PARADOX AS THEOREM

Arthur Prior advocated something along the lines that I would urge.¹⁰ (Prior observed, rightly, that Cohen had not really seen the paradoxical nature of the situation.) Prior ended up accepting what appears to be some extreme oddities in the theory of propositions. This is the price one will have to pay on the proposed alternative to ramification. Before I state the alternative, I want to "soften up" the reader with a simple arithmetical theorem about functions and functionals. Let the individual variables range over the natural numbers and let the function variables range over functions from natural numbers to natural numbers. Define the functional: $\Pi(f, g) = \begin{cases} 0, & \text{if for every } x \text{ such that } f(x) = 0, g(x) = 0. \\ 1, & \text{Otherwise.} \end{cases}$ And define $\nu(x) = \begin{cases} 0, & \text{if } x \neq 0 \\ 1, & \text{if } x = 0 \end{cases}$. We might call ν "the switch function." It just yields 1 for argument 0 and 0 for everything else.

Theorem¹¹ $\forall f[\forall x[f(x) = 0 \rightarrow \Pi(f, \nu) = x] \rightarrow \forall x \sim [f(x) = 0]]$.

That is, "If for every argument of a function for which it has the value 0, Π applied to it together with the 'switch function' ν is equal to that argument, then the function does not have the value 0 for *any* argument." The proof is quite elementary, but we include it to ward off any possible future doubts about its validity.

8. "Self-Reference in Philosophy," in *Symbolic Logic: An Introduction* (New York: Ronald Press, 1952), appendix.

9. "Russell's Mathematical Logic," in *The Philosophy of Bertrand Russell*, ed. Paul A. Schilpp (New York: Tudor, 1944).

10. "Epimenides the Cretan," *Journal of Symbolic Logic* 23 (1958): 261–266.

11. The form (but not the matter) of this theorem was suggested to me by Theorem 69 of Church's "A Revised Formulation of the Logic of Sense and Denotation. Alternative (1)," *Noûs* 27 (1993): 141–157.

Proof: Let ν and Π be as defined and suppose, for a reductio, that there is some function f such that $\forall x[f(x) = 0 \rightarrow \Pi(f, \nu) = x]$, but that $\exists x[f(x) = 0]$. Call such an x , x_0 , so that $f(x_0) = 0$. Then by the condition stated, we must have that $f(x_0) = 0 \rightarrow \Pi(f, \nu) = x_0$, whence $\Pi(f, \nu) = x_0$. It is clear from the definition of Π , that $x_0 = 0$ or $x_0 = 1$.

If $x_0 = 0$, then $\Pi(f, \nu) = 0$, so that, by the definition of Π , since $f(x_0) = 0$, we must have that $\nu(x_0) = 0$. Hence, by the definition of ν , $x_0 \neq 0$. Contradiction.

If $x_0 = 1$, then $\Pi(f, \nu) = 1$. So, by the definition of Π , for some y , $f(y) = 0$, but $\nu(y) \neq 0$. Call one such, y_0 , so that $f(y_0) = 0$ and $\nu(y_0) \neq 0$. By the definition of ν , we must have that $y_0 = 0$. But we have assumed that $\forall x[f(x) = 0 \rightarrow \Pi(f, \nu) = x]$. So we have $f(y_0) = 0 \rightarrow \Pi(f, \nu) = y_0$. Thus $\Pi(f, \nu) = y_0$, i.e. $\Pi(f, \nu) = 0$. Again, contradiction. ■

Now it is not difficult to show that this theorem implies (and, in fact, is equivalent to):

$$\forall f \forall x \{ [x = \Pi(f, \nu)] \rightarrow \sim \forall y (f(y) = 0 \leftrightarrow x = y) \}.$$

If you think about it, the proof given above works no matter what domain we take, as long as we pick 0 and 1 to belong to that domain and to be distinct. Suppose we take the domain to be a set of propositions and we take 0 to be some necessarily true proposition and 1 to be some necessarily false proposition, call them **t** and **f**, respectively.

Let us take such a domain and use variables that suggest that choice. Rewriting:

$$\forall \phi \forall p \{ [p = \Pi(\phi, \nu)] \rightarrow \sim \forall s (\phi(s) = \mathbf{t} \leftrightarrow p = s) \}$$

Let us rewrite this using ' $\neg$ ' instead of ' ν ', and use ' $(r)(\phi(r) \supset \neg r)$ ' as a long-winded way of writing ' $\Pi(\phi, \neg)$ '. Finally, let us take the liberty of writing $\phi(s)$ for $\phi(s) = \mathbf{t}$. Then it looks like this:

$$\forall \phi \forall p \{ [p = (r)(\phi(r) \supset \neg r)] \rightarrow \sim \forall s (\phi(s) \leftrightarrow p = s) \}.$$

If we choose ϕ to be the characteristic function of some set of propositions (where **t** and **f** are used as the values of characteristic functions), call it Q , we get:

$$(E') \forall p \{ [p = (r)(Q(r) \supset \neg r)] \rightarrow \sim \forall s (Q(s) \leftrightarrow p = s) \}.$$

This seems to be no more or less paradoxical than the "Epimenides' Theorem":

$$(E) \forall p \{ [p = \forall s (Qs \rightarrow \sim s)] \rightarrow \sim \forall s (Qs \leftrightarrow p = s) \}.$$

This is in fact a theorem of the natural formulation of the extended propositional calculus described above. If modality is present, then we may necessitate (E). Given the Converse Barcan Formula (which is difficult to fault in the present instance), we can obtain:

$$(E\Box) \forall p \Box \{ [p = \forall s (Qs \rightarrow \sim s)] \rightarrow \sim \forall s (Qs \leftrightarrow p = s) \}$$

“Every proposition is necessarily such that if it is the proposition that all Q s are untrue, then it is not the *unique* Q proposition.” Without ramification we will have (S) above. Distributing the necessity in (E $\Box$) and using some elementary quantification theory, one obtains the *theorem*:

$$(-A) \exists p \Box \sim \forall s (Qs \leftrightarrow p = s).$$

That is, there is a proposition that is necessarily not the unique Q . Notice that there has been no assumption, explicit or implicit, that propositions are to be construed in the way favored by possible world semantics. This theorem, “anti-(A),” contradicts (A) above.

There is no doubt that for particular choices of Q the Epimenides Theorem and its apparent consequence anti-(A) are extremely odd. Epimenides just *cannot* assert that all things asserted by Epimenides are untrue and nothing else, although it looks for all the world as if he can. Similar oddities appear when the “policeman-prisoner” example above is formalized along these lines. Here the conclusion is the formal version of this: If the policeman asserts exactly one thing and the prisoner deposes that something the policeman asserts is correct, then it cannot be both that the policeman asserted only that everything the prisoner deposed was untrue and that the prisoner asserted only that something the policeman said was true. Both can utter words that would appear to say these things. But they cannot. So, this line does not preserve the appearances. There *is* such a thing as the proposition that all propositions asserted by the prisoner are untrue and a proposition to the effect that something asserted by the policeman is true—but these cannot be the sole things asserted by the policeman and the prisoner, respectively.

The ramification approach really does not preserve the appearances either. Epimenides, the policeman, and the prisoner cannot have been saying what they appear to have been saying. But with ramification there is *something like* that they intended that might in fact have been asserted. That is a sort of a half-plus for ramification.

In summary: Kaplan has shown that possible world semantics cannot preserve (A). He has urged that logic should be neutral on metaphysical questions, but this line seems very difficult to follow in general. The additional paradox does not directly turn on possible world semantics but is the Epimenides come back to haunt us. Kaplan proposes ramification to thwart the paradox and to preserve a ramified version of (A). I have urged a different course. It has some advantages over ramification. But the cost is undeniable.¹²

12. This essay is a slightly revised version of a paper presented at the San Marino conference honoring David Kaplan. David Kaplan is my teacher and friend, and my criticism of his work here should not be construed as suggesting that this or any other of his interesting contributions to philosophy are not worthy of extreme admiration—as I in fact believe that they are.

CHAPTER 7

Possible Worlds Semantics and the Liar

Reflections on a Problem Posed by Kaplan

Sten Lindström

In this essay I discuss a paradox, due to David Kaplan, that in his view threatens the use of possible worlds semantics as a model-theoretic framework for intensional logic.¹ Kaplan's paradox starts out from an intuitively reasonable principle that I refer to as the *principle of plenitude*. From this principle he derives a contradiction in what he calls *naive possible world theory*. Kaplan's metatheoretic argument can be restated in the modal object language as an intensional version of the *Liar paradox*.

To dissolve the paradox, Kaplan favors a ramified theory of propositions, along the lines of Russell's ramified theory of types. According to this approach: (i) The propositions are divided into orders 0, 1, ...; (ii) a propositional quantifier never ranges over the totality of all propositions but only over the propositions of some given order; and (iii) a proposition that involves quantification over the propositions of a given order will itself be of higher order than the propositions in the domain of quantification. Accordingly, a sentence can never express a proposition that is itself within the range of one of the propositional quantifiers in the sentence. This approach avoids the intensional Liar paradox, but its implementation involves severe complications both in the syntax and in the model theory of intensional logic.²

Here I shall attempt an alternative, less drastic, modification of the standard possible worlds methodology than the one favored by Kaplan.

1. The fullest and most recent presentation of the argument is Kaplan (1994). Kaplan, who has been communicating the argument since the mid-1970s, presented it as part of his John Locke lectures in Oxford during the spring of 1980 and at the *Seventh International Congress of Logic, Methodology and Philosophy of Science* in Salzburg, Austria, July 11–16, 1983 (an abstract of the argument appeared in *Abstracts of Sections 5 and 12: Seventh International Congress of Logic, Methodology and Philosophy of Science*). It has been reported and discussed in Cresswell (1980); Davies (1981); Lewis (1986); and Jubien (1988).

2. It may be instructive to consider Russell's informal explanation in Whitehead and Russell (1910–1913), vol. 1, p. 62, of the resolution of the Liar paradox within the ramified theory of types:

The idea is to regard sentences that involve propositional quantifiers, like the sentence

(A) All propositions contemplated by Epimenides are false,

as being, in a sense, indexical: One and the same sentence can express different propositions when used in different possible worlds. The context of use fixes the range of the propositional quantifiers in the sentence and thereby the proposition expressed by the sentence. Using this approach, I try to show that the intensional Liar paradox can be defused and no longer poses a threat to possible worlds semantics. This means that the principle of plenitude can consistently be upheld in the modified semantics.

1. BACKGROUND: INTENSIONAL LOGIC AND PARADOX

General intensional logic—the logical study of so-called oblique or intensional constructions—is an area that is thriving on paradoxes and puzzles but is also haunted by them. There are, of course, the puzzles of intensionality that started the whole enterprise and constitute its *raison d'être*: Frege's puzzle about the information value of true identity statements, Russell's puzzle concerning the author of *Waverley*, puzzles by Quine and others about quantifying in and *de re* constructions, Kripke's puzzle about belief, and so on. It is probably safe to say that even today there is no consensus among logicians concerning the correct treatment and diagnosis of these puzzles.

The various approaches to intensional logic that were proposed to take care of the initial puzzles soon turned out to create paradoxes and foundational difficulties of their own. These approaches can be divided into direct-discourse and indirect-discourse ones. According to the *direct-discourse treatments*, oblique constructions in ordinary language (like 'John knows *that A*', 'It is necessary *that A*') are represented in the regimented languages of logic by direct-discourse analogues ('John knows *A*', '*A* is a necessary truth', where *A* is a *name* rather than a formula). Necessity, for instance, is represented not by a sentential *operator* that takes a sentence (or a formula) *A* and yields a new sentence (or formula) $\Box A$ but by a *predicate* *Nec*, which when appended to a *name A* yields a sentence *Nec(A)*. Direct-discourse approaches come in two varieties, according to whether the name *A* is taken to designate the expression *A* itself or the sense (or intension) associated

When a man says "I am lying", we may interpret his statement as: "There is a proposition which I am affirming and which is false". . . . But we saw that the word "false" is ambiguous, and that, in order to make it unambiguous, we must specify the order of falsehood, or what comes to the same thing, the order of the proposition to which falsehood is ascribed. We saw also that, if *p* is a proposition of the *n*th order, a proposition in which *p* occurs as an apparent variable is not of the *n*th order but of a higher order. Hence the kind of truth or falsehood which can belong to the statement "there is a proposition *p* which I am affirming and which has falsehood of the *n*th order" is a truth or falsehood of a order higher than the *n*th. Hence, the statement of Epimenides does not fall within its own scope, and therefore no contradiction emerges.

with *A*. The first, the *syntactic approach*, has been advocated by Quine; the second, the *intensional direct-discourse approach*, was adopted by Church in his logic of sense and denotation.³

The direct-discourse approaches view the intensionality of natural language, that is, its failure of extensionality, as a surface phenomenon: The underlying logic and its semantics is taken to be fully extensional. The *indirect-discourse approaches*, on the other hand, work with logical languages where oblique constructions are represented by intensional operators.

The direct-discourse approaches are, however, threatened by (self-referential) paradoxes. Montague (1963) showed that the syntactic treatment of necessity as a predicate of sentences in the object language leads to inconsistency. Kaplan and Montague (1960) contains a similar result for knowledge ("the Knower paradox"). These results generalize Tarski's theorem on the undefinability of the truth-predicate in a sufficiently rich object language.

Also the intensional direct-discourse approach is threatened by self-referential paradoxes. Church's logic of sense and denotation is not directly concerned with linguistic expressions and their senses and denotations, but rather with the language-independent concept relation Δ between concepts (Fregean senses) and the entities they determine (denotations). The more finely the concepts are individuated, the more closely will the abstract theory of concepts and their objects resemble the more concrete theory of linguistic expressions and their denotations, with Church's concept relation playing a role similar to the one played by the denotation predicate of semantics. Consequently, antinomies analogous to the semantic antinomies may arise for formulations of the logic of sense and denotation along the lines of Church's so-called Alternatives (0) or (1), where logically equivalent expressions are allowed to have distinct senses. Indeed, as Myhill (1958) points out, Church's Alternative (0) is threatened by the antinomy described by Russell in *The Principles of Mathematics*, Appendix B, p. 527—the so-called *Russell-Myhill Paradox*. Myhill argues as follows:

The system Church designed to embody Alternative 0 is formally inconsistent! For if *f* and *g* are two distinct sets of propositions, the proposition that every proposition belongs to *f* is distinct, in virtue of the "principle of

3. See also Russell (1908) and Church (1976) for discussions of the so-called semantical paradoxes within the context of the ramified theory of types.

Cf. Quine (1953) and Church (1951, 1974). The distinction between the direct-discourse approach and the indirect-discourse approach to intensional logic was clearly stated in Kaplan (1964). In that work, Kaplan is concerned with developing the direct-discourse approach of Alonzo Church and providing it with a model-theoretic semantics of the possible worlds variety. The model-theoretic study of Church's logic of sense and denotation within possible worlds semantics is pursued further in Kaplan (1975) and Parsons (1982). For an early discussion of different methods of semantic analysis in intensional logic, including Church's, see Carnap (1947).

maximum distinction" characteristic of Alternative 0, from the proposition that every proposition belongs to g . There is thus established a one-to-one mapping of sets of propositions into propositions, in violation of Cantor's theorem. There is no difficulty in formalizing this argument as a derivation from Church's axioms.

The remedy proposed by Church (1974) and Anderson (1980) is to split up the concept relation $\Delta(x, y)$, holding between a concept x and an object y when the former determines the latter, into an infinite sequence Δ^m of relations satisfying the condition:

$$\forall x \forall y [\Delta^m(x, y) \rightarrow \Delta^{m+1}(x, y)],$$

thereby creating a *ramified* intensional type theory rather than a simple one. Anderson (1980) presented axioms for such a system and showed them to be consistent by constructing a model. The resulting system, however, is extremely complicated.

As indicated, the direct-discourse approach has faced immense difficulties, and the systems that have been created have been so complicated that to most researchers the approach has hardly seemed worth pursuing. In comparison, the indirect-discourse approach of modal logic and Montague's intensional logic has been greatly successful. Oblique constructions of natural language are here represented by intensional operators, and the formal language is provided with a model-theoretic semantics of the possible worlds variety. Although there are problems with this approach, for example, the problem of logical omniscience—and difficulties of an interpretational nature concerning the concept of a possible world—it seems that at least the self-referential paradoxes have been avoided. But is this really so? In this essay I discuss an argument due to David Kaplan according to which possible worlds semantics may, after all, also be threatened by the Liar paradox and its relatives. If Kaplan is right, the indirect-discourse approach and possible worlds semantics are faced with problems analogous to those facing the direct-discourse approach.

2. KAPLAN'S PROBLEM

There is, according to Kaplan, a serious foundational problem in what might be called *naïve possible world theory*—a problem that is analogous to the paradoxes of naïve set theory and that threatens the use of the theory as a basis for the model theory of intensional languages. Kaplan's argument starts out from the following observation: Consider some thinker and some time t . Then the following principle appears to be logically consistent:

(P) For any proposition p , it is possible that the thinker entertains the proposition p at time t and that p is the only proposition that he entertains at time t .

It may be that the principle (P) is ruled out for metaphysical reasons: Perhaps there must be, for any thinker, propositions that are not possible contents of thought for that individual. However, Kaplan argues, logic alone should not exclude the situation envisaged in (P). In other words, intensional logic should not rule out the existence of a property Q of propositions satisfying the principle:

$$(P) \forall p \Diamond (Qp \wedge \forall q (Qq \rightarrow q = p)),$$

where the quantifiers are thought of as ranging over the collection of (absolutely) all propositions.⁴ I shall refer to this principle as Kaplan's *principle of plenitude*.

Now, Kaplan's argument shows that the principle of plenitude is incompatible with assumptions commonly made in possible worlds semantics. Here is how the argument goes:

- (i) There is a set W of possible worlds and a set $Prop$ of propositions.
- (ii) There is, for every subset X of W , a corresponding proposition $[X]$ that is true with respect to any possible world w just in case w belongs to X .
- (iii) Different sets of possible worlds correspond to different propositions. So $card(Prop) \geq card(\wp(W)) > card(W)$.
- (iv) By the principle of plenitude, there is a property Q of propositions such that: for each proposition p , it is possible that p should have had the property Q and that p should have been the only proposition having that property.
- (v) Hence, for each proposition p , there is a possible world w_p in which it is true that p has the property Q and that p is the only proposition having the property Q .
- (vi) Clearly, if p and q are distinct propositions, then the possible worlds w_p and w_q must also be distinct. That is, the function mapping p to w_p is one-to-one.
- (vii) From (vi) follows that $card(W) \geq card(Prop)$, contrary to (iii).

So this is Kaplan's paradox: According to possible worlds semantics, there are at least as many propositions as there are sets of possible worlds. So, by Cantor's theorem, there are more propositions than there are

4. The reader might wonder about the exact formalization of the principle of plenitude. Perhaps, we should rather write it as:

$$(P') \forall p \Diamond \exists q (q = p \wedge Qq \wedge \forall r (Qr \rightarrow r = q))$$

or as

$$(P'') \forall p \Diamond \forall q (Qq \leftrightarrow q = p).$$

Given a possible worlds semantics where the quantifiers range over a constant domain containing all propositions, the three formulations are equivalent. But what about the case when we allow for the domain of quantification to vary from one world to another? Even then, the three formulations are equivalent provided that the operator Q satisfies the requirement:

$$\forall p \Box (Qp \rightarrow \exists q (q = p \wedge Qq)).$$

In its absence, (P') will be our official formalization of the principle of plenitude.

worlds. On the other hand, an adequate intensional logic should be compatible with the principle of plenitude. But according to this principle, there are at least as many possible worlds as there are propositions. Hence, possible worlds semantics cannot meet the demands of intensional logic.

It should perhaps be pointed out that Kaplan's problem has nothing to do with the assumption, standardly made within possible worlds semantics, that propositions with the same intension, that is, that are true in exactly the same worlds, are identical. This so-called *principle of intensionality* was not assumed above and does not play any role in Kaplan's argument. In other words, the argument is relevant also for conceptions of propositions that take them to be more fine-grained than Carnapian intensions (i.e., sets of possible worlds).

3. THE PRINCIPLE OF PLENITUDE AND THE LIAR

One obvious reaction to Kaplan's problem is to question the assumption that for every set of possible worlds there exists a proposition having that set as its intension. In his *On the Plurality of Worlds* (1986, sec. 2, pp. 104–108), David Lewis suggests this as the solution to the paradox: Not all propositions are possible contents of thought for a given agent; therefore, the class of propositions to which the principle of plenitude is applicable cannot correspond to the entire power set of the set of worlds.

This does not seem to be a fully adequate reaction to Kaplan's problem. First of all, as Kaplan stresses, even if, for deep metaphysical reasons, the set of all possible contents of thought cannot exhaust the power set of W , it does not seem reasonable to exclude the possibility that it does when trying to devise a model theory for intensional languages. But there is an even stronger reason why the above reaction is inadequate: Even if we give up the assumption that there are propositions corresponding to all sets of worlds, paradox still threatens. As Kaplan (1994) himself points out, the principle (P) gives rise to an intensional version of the Liar paradox. Consider, namely, the sentence:

- (A) No proposition having the property Q is true.

We make the intuitively reasonable assumption that there is a proposition, let us call it λ , that the sentence A expresses. Then,

- (i) It is necessary that: λ is true iff no proposition having the property Q is true.

However, according to the principle of plenitude:

- (ii) It is possible that: λ has the property Q and λ is the only proposition having the property Q .

From (i) and (ii) we get:

- (iii) It is possible that: λ is the one and only proposition having property Q and (λ is true iff no proposition having the property Q is true).

Finally, (iii) yields:

- (iv) It is possible that: λ is true iff λ is not true,

which is an obvious contradiction.

In order to formalize the above argument, we introduce a suitable formal language L . The *symbols* of L are (i) a denumerable sequence of propositional variables $p_1, p_2, \dots$; (ii) the Boolean connectives $\perp$ and $\rightarrow$; (iii) the identity connective $=$; (iv) the universal quantifier symbol $\forall$; (v) the logical propositional operator $\Box$ (the necessity symbol); (vi) a non-logical propositional operator Q ; (vii) parentheses: '(' and ')'. We use the letters ' p ', ' q ', and ' r ' to range over propositional variables in L . The symbols ' $\neg$ ', ' $\wedge$ ', ' $\vee$ ', ' $\leftrightarrow$ ', ' $\exists$ ', ' $\Diamond$ ', are introduced as abbreviations in the usual way.

The set of *formulas* of L is the smallest set Φ that is closed under the conditions:

- (i) all propositional variables of L belong to Φ ;
- (ii) $\perp \in \Phi$;
- (iii) if $A, B \in \Phi$, then $(A \rightarrow B)$ and $(A = B) \in \Phi$;
- (iv) if $A \in \Phi$, then $\Box A \in \Phi$ and $QA \in \Phi$;
- (v) for any propositional variable p and $A \in \Phi$, $\forall p A \in \Phi$.

The notions of a variable occurring *free* in a formula is defined in the usual way. A *sentence* of L is a formula in which no variables occur free.

We now turn to the task of formalizing the above argument as a natural deduction proof in L . We symbolize the Liar sentence as $\forall p(Qp \rightarrow \neg p)$, or in primitive notation, $\forall p(Qp \rightarrow (p \rightarrow \perp))$. Below we use ' Λ ' as an abbreviation for this sentence.

We start the argument by assuming that (1) there is a proposition expressed by Λ ; and (2) the principle of plenitude holds:

{1}	(1)	$\exists r(r = \Lambda)$	Premise
{2}	(2)	$\forall p \Diamond \exists q(q = p \wedge Qq \wedge \forall r(Qr \rightarrow r = q))$	Premise

Next, we go through the Liar argument itself:

{3}	(3)	$\exists q(q = \Lambda \wedge Qq \wedge \forall r(Qr \rightarrow r = q))$	Premise
{3}	(4)	$Q\Lambda$	3 by predicate logic
{3}	(5)	$\forall p(Qp \rightarrow p = \Lambda)$	3 by predicate logic
{6}	(6)	Λ (i.e., $\forall p(Qp \rightarrow (p \rightarrow \perp))$)	Premise
{3, 6}	(7)	$\Lambda \rightarrow \perp$	4, 6 by predicate logic
{3, 6}	(8)	$\perp$	6, 7 modus ponens ($\rightarrow E$)
{3}	(9)	$\Lambda \rightarrow \perp$	6–8 by $\rightarrow I$
{3}	(10)	$Q\Lambda \rightarrow (\Lambda \rightarrow \perp)$	9 using $\rightarrow I$
{3}	(11)	Λ	5, 10 by predicate logic

{3}	(12)	$\perp$	9, 11 modus ponens
{}	(13)	$\exists q(q = A \wedge Qq \wedge \forall r(Qr \rightarrow r = q)) \rightarrow \perp$	3, 12 $\rightarrow$ I

That is, from the assumption:

The one and only proposition having the property Q is the proposition that all propositions having the property Q are false

follows logically the necessarily false proposition $\perp$.

So far we have used only nonmodal predicate logic (as a matter of fact, we have used only the very weak system of *minimal logic* (cf. Prawitz 1965)). Assuming that the modal logic for $\Box$ is normal, we have the inference rule:

If $A \rightarrow B$ is provable, then $\Diamond A \rightarrow \Diamond B$ is also provable.

By means of this rule, we infer from (13):

{} (14) $\Diamond \exists q(q = A \wedge Qq \wedge \forall r(Qr \rightarrow r = q)) \rightarrow \Diamond \perp$.

But, (1) and (2) yield:

{1, 2} (15) $\Diamond \exists q(q = A \wedge Qq \wedge \forall r(Qr \rightarrow r = q))$.

So, finally we have:

{1, 2} (16) $\Diamond \perp$

contrary to the assumption that the modal logic is normal.

In light of this argument it seems that we have to give up either the principle of plenitude or the intuitively reasonable assumption that the Liar sentence A expresses an (actually existing) proposition. The argument does not explicitly involve possible worlds or possible worlds semantics. Instead, it uses simple and seemingly natural principles of modal logic. At first sight, the argument casts considerable doubt on the principle of plenitude. As a matter of fact, it can easily be formalized within Montague's intensional logic *IL* or Gallin's higher-order modal logic (cf. Montague 1970, 1974; Gallin 1975). Within Montague's intensional logic, it is provable that every sentence A expresses a proposition. Hence, we can formally prove within *IL* that the principle of plenitude is false.

But, as the saying goes, "One person's modus ponens is another person's modus tollens." So, this result can as well be taken as an indication that systems like Montague's *IL* are unreasonably strong. As a matter of fact, we think that this is the right moral to draw from Kaplan's argument. The argument is not so much an objection to possible worlds semantics as a warning against an uncritical application of the possible worlds methodology that might lead logicians to construct systems of intensional logic in which intuitively consistent principles, like the principle of plenitude, are provably inconsistent.

4. THE INTENSIONAL LIAR AS A DIAGONAL ARGUMENT

Suppose that we distinguish between the set *Prop* of all propositions and the subset *U* of *Prop* over which the propositional quantifiers of *L* range. We speak of *U* as the domain of quantification. The Liar proposition for *U* is the proposition λ_U expressed by the sentence:

$$(A) \quad \forall p(Qp \rightarrow \neg p),$$

when we take the universal quantifier in *A* to range over *U*. We can now view the Liar argument as a proof of the following:

Theorem If the domain of quantification *U* satisfies the principle of plenitude, then the Liar proposition for *U* is not a member of *U*.

Proof: Suppose that *U* satisfies plenitude, i.e.,

(1) for every $p \in U$, there exists a $w \in W$ such that *Qp* is true in *w* and for every $q \in U$, if *Qq* is true in *w*, then $q = p$.

For any world $w \in W$, we have that:

(2) λ_U is true in *w* iff for every $p \in U$, if *Qp* is true in *w*, then *p* is not true in *w*.

Suppose now that $\lambda_U \in U$. Then, by (1),

(3) there exists $w \in W$ such that $Q\lambda_U$ is true in *w* and for every $q \in U$, if *Qq* is true in *w*, then $q = \lambda_U$.

However, by the usual Liar argument, we get for the world *w* that exists according to (3) that:

(4) λ_U is true in *w* iff λ_U is not true in *w*.

Hence, by reductio, $\lambda_U \notin U$.

Corollary If the domain of quantification *U* satisfies the principle of plenitude, then it cannot be the whole set *Prop*.

5. A RESOLUTION OF KAPLAN'S PARADOX SUGGESTED BY THE DIAGONAL ARGUMENT

Viewing the Liar paradox as a diagonal argument suggests a way out of Kaplan's problem. In the diagonal argument above, we interpreted the Liar sentence *A* relative to a parameter, namely, the domain *U* of propositions over which we quantify. The idea is that the sentence *A* will express different propositions for different values of the parameter *U*.

Suppose now that this parameter is fixed by context, that is, for every context of use *c*, there is a domain U_c of all the propositions that are *available* in *c*. Propositions that lie outside of U_c , we think of as *unavailable* in *c*: They cannot be quantified over from the perspective of *c*. When from the perspective *c*, we speak of "all propositions," we always refer to the totality U_c of all propositions that are available in *c*. This is assumed to

be the case also when we speak of counterfactual situations (possible worlds). Hence, the domain of quantification is assumed to be rigidly fixed by context. On this approach, the Liar sentence:

(A) All propositions contemplated by Epimenides are false

will in general express different propositions relative to different contexts of use c .

For any context of use c , A expresses relative to c the proposition λ_c such that for any world w ,

λ_c is true at w iff for all propositions p in U_c , if p is contemplated by Epimenides in w , then p is false (not true) in w .

Given that the proposition expressed by the Liar sentence has been determined as λ_c by the actual context c , we can evaluate the proposition λ_c at various points of evaluation w . Letting the possible worlds serve the double roles of *contexts of use* as well as *points of evaluation*, we can in particular evaluate the proposition λ_c at the point c itself (the actual world). We write $\models_{u,v} A$, for a formula A taken in the context u being true at the point of evaluation v . In the next section, we shall give a rigorous definition of this notion. Let us for the time being leave it at the intuitive level.

Returning now to the Liar, we ask ourselves whether it is possible for λ_c to be the one and only proposition contemplated by Epimenides in the actual world c . Let us say that a sentence A is *true in a context c* (in symbols, $\models_c A$) iff $\models_{c,c} A$, that is, iff A taken in the context c is true at c . A is false in the context c iff $\neg A$ is true in the context c . Thus, A is true in the context c just in case the proposition expressed by A in c is true at c . In particular, the Liar sentence is true in the context c iff the proposition λ_c is true at c . Hence, we have:

$\models_c \lambda$ iff for all propositions p in U_c , if p is contemplated by Epimenides in c , then p is not true in c .

Suppose now that the following sentence is true in c ,

(1) $\exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \lambda)$

where we interpret ' Qp ' as "Epimenides contemplates the proposition p ."

Then it is true (in the actual world) that:

The proposition that all propositions contemplated by Epimenides are false is the one and only (available) proposition contemplated by Epimenides.

But this means that the proposition λ_c is in U_c and is the unique proposition contemplated by Epimenides in c . By the diagonal argument, again, we get:

$\models_c \lambda$ iff it is not the case that $\models_c \lambda$.

We conclude that the sentence (1) cannot be true in the actual world. In other words, the negation of (1):

$$(2) \neg \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \Lambda)$$

is *logically true* in the sense of expressing a true proposition in every context (in every model). Regardless of which world is the actual one, (2) expresses a true proposition. Hence, we can say that the *sentence* (2) is a priori true.

(2) leaves the possibility that λ_c is the one and only proposition p in *Prop* that Epimenides contemplates in the actual world c , *provided that* λ_c is not a member of U_c . In other words, it is still possible that the proposition expressed by

$$(\Lambda) \forall p(Qp \rightarrow \neg p)$$

is the one and only proposition contemplated by Epimenides in the actual world. This possibility obtains, as long as this proposition is not within the range assigned to the quantifier in Λ in the actual world.

As Myhill (1979, 83) points out, the intensional Liar that we have considered should be distinguished from those so-called semantic paradoxes that concern linguistic entities like sentences and predicates rather than propositions and concepts:

Actually what are called semantic paradoxes are of two kinds, exemplified by the Epimenides and the heterological paradox, respectively. The first kind does not involve any mention of expressions, but only of propositions; it does not depend on syntax, but (in the case of Epimenides) on the pragmatic relation of assertion. The other kind includes, besides the heterological paradox, the paradox of the least integer not nameable in fewer than nineteen syllables; of the least undefinable ordinal; and of Richard's. It *also* includes 'This sentence is false' which can be rigorously presented using Gödel's diagonalization technique, and which is concerned with falsehoods of *sentences* not *propositions* (and therefore is distinct from the Epimenides with which it is frequently confused).

I am inclined to look upon the intensional Liar ("Epimenides") on the analogy of the pseudo-paradox of the Barber.⁵ We escape the Barber by denying the existence of any such person (i.e., a village barber that shaves all and only those inhabitants of his village that do not shave themselves). The assumption that there is such a person is simply logically inconsistent. Similarly, it seems that we can avoid the intensional Liar by viewing it as a *reductio ad absurdum* of the assumption that there is a person who actually contemplates (or asserts) one and only one proposition in the domain U_c of available propositions, namely, the proposition that all

5. Myhill (1979) is of the opposite opinion: He seems to view the intensional Liar as a genuine antinomy that requires a division of propositions into types along the lines of Russell's Ramified Theory of Types.

propositions (in U_c) that he contemplates are false. On this view, the assumption of the intensional Liar:

$$(1) \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \Lambda)$$

is simply inconsistent.

Suppose now that one and only one sentence is engraved on Epimenides' tomb, namely:

$$(\Lambda) \text{ All propositions expressed by sentences engraved on Epimenides' tomb are false}^6$$

and let λ_c be the proposition that is actually expressed by Λ . Then, it seems reasonable to assume that λ_c is the only proposition satisfying the condition of being expressed by a sentence engraved on Epimenides' tomb. By the diagonal argument, however, we can conclude that:

$$\lambda_c \notin U_c,$$

that is, the proposition expressed by Λ in the actual world cannot be in the domain of propositions over which the propositional quantifier in Λ ranges, when Λ is interpreted in the actual world. In other words, if the sentence on Epimenides' tomb expresses a (unique) proposition at all, we can be sure that this proposition is not within the range of the propositional quantifiers that occurs in it. Therefore, (1) is not true, and the Liar paradox is avoided.

But where does this leave Kaplan's principle of plenitude? According to standard modal logic, the logical falsity of (1) implies the logical truth of

$$(2) \neg \Diamond \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \Lambda),$$

which apparently contradicts the principle of plenitude:

$$\forall r \Diamond \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = r).$$

Does this mean that the sentence (1) is necessarily false? Couldn't there have been a counterfactual situation w in which λ_c is the one and only proposition that is contemplated by Epimenides? According to the approach envisaged here, such a situation is indeed possible. Consider, for example, a world w whose set of available propositions U_w is a proper superset of the set U_c of actually available propositions. Then, λ_c could very well be the one and only proposition that Epimenides contemplated in w . Thus, we could have:

$$\models_{c,w} \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \Lambda)$$

from which we could infer:

$$\models_{c,c} \Diamond \exists p(Qp \wedge \forall q(Qq \rightarrow q = p) \wedge p = \Lambda).$$

6. This variant formulation of the paradox is, of course, due to Kaplan (1994).

The explanation for this possibility is that λ_c is not a Liar proposition with respect to the situation w : It is not necessarily equivalent to the proposition λ_w saying that for all propositions in U_w , if Epimenides contemplates p , then p is false. It would indeed have been impossible for λ_w —the proposition expressed by Λ relative to w —to be the one and only proposition contemplated by Epimenides in w . In order to express this impossibility, we need to extend the language L with the *index operator* $\dagger$ having the evaluation clause:

$$\models_{c,u} \dagger A \text{ iff } \models_{u,u} A.$$

Thus, the index operator $\dagger$ takes the current point of evaluation and makes it the point of reference.⁷ In terms of $\dagger$, we can now express:

$$\neg \Diamond \dagger \exists p (Qp \wedge \forall q (Qq \rightarrow q = p) \wedge p = \Lambda).$$

This says that there cannot be a world w such that $\lambda_w \in U_w$ and λ_w is the one and only proposition contemplated by Epimenides in w .

It should be clear from the above discussion that given this account, there is nothing that stops the principle of plenitude:

$$\forall p \Diamond \exists q (q = p \wedge Qq \wedge \forall r (Qr \rightarrow r = q))$$

from holding in the actual world.

APPENDIX: THE FORMAL SEMANTICS

For the sake of completeness, we write up the formal semantics for L (with the added operator $\dagger$) that was outlined informally in the previous section.

We define a *frame* to be a sequence:

$$F = \langle W, U \rangle$$

satisfying the conditions:

- (i) W is a nonempty set (referred to as the set of *possible worlds*).
- (ii) U is a function that assigns to each $w \in W$ a subset U_w of $\wp(W)$.

Subsets of W are called *propositions*. For each $w \in W$, we say that U_w is the set of propositions that are *available* in w .

A *model* for the language L is a sequence $M = \langle W, U, V \rangle$, where

- (i) $F = \langle W, U \rangle$ is a frame; and
- (ii) V is a function that assigns to the non-logical operator symbol Q , an operation on propositions $V(Q): \wp(W) \rightarrow \wp(W)$.

$M = \langle W, U, V \rangle$ is said to be a model based on $F = \langle W, U \rangle$. An *assignment* (in M) is a function g that assigns a proposition $g(p) \in \wp(W)$ to every propositional variable p in L .

7. The $\dagger$ -operator is discussed in Lewis (1973), sec. 2.8. See also Segerberg (1973).

Given a model $M = \langle W, U, V \rangle$ and a possible world $w \in W$ (serving as context), we define for every formula A of L and every assignment g , the proposition $|A[g]|_w$ expressed by A at w , with respect to M, g , as follows:

- (i) if p is a propositional variable, then $|p[g]|_w = g(p)$;
- (ii) $|Q(A)[g]|_w = V(Q)(|A[g]|_w)$;
- (iii) $|(A = B)[g]|_w = W$, if $|A[g]|_w = |B[g]|_w$; and $|(A = B)[g]|_w = \emptyset$, otherwise;
- (iv) $|\perp[g]|_w = \emptyset$;
- (v) $|(A \rightarrow B)[g]|_w = (W - |A[g]|_w) \cup |B[g]|_w$;
- (vi) $|\forall p A(p)[g]|_w = \{u \in W : \text{for all } a \in U_w, u \in |A(p)[g(a/p)]|_w\}$, where $g(a/p)$ is the assignment that is obtained from g by assigning the proposition a to the propositional variable p and letting $g(a/p)(q) = g(q)$ for all other propositional variables q .
- (vii) $|\Box A[g]|_w = W$, if $|A[g]|_w = W$; and $|\Box A[g]|_w = \emptyset$, otherwise;
- (viii) $|\dagger A[g]|_w = \{u \in W, u \in |A[g]|_u\}$.

Observe that the only role for the context parameter w is to fix the domain U_w over which the propositional variables range. Once the domain of quantification is fixed by context, it is taken to be constant for all points of evaluation.

For each formula A , model M , pair of worlds u, v in M , and assignment g into M , we define:

$$M \models_{u,v} A[g] \text{ iff } v \in |A[g]|_u.$$

We say that A is *true at v with reference to M, g and u* iff $M \models_{u,v} A[g]$. We have:

- Not: $M \models_{u,v} \perp[g]$;
- $M \models_{u,v} (A \rightarrow B)[g]$ iff either (not: $M \models_{u,v} A$) or $M \models_{u,v} B[g]$;
- $M \models_{u,v} \Box A[g]$ iff for all $w \in W$, $M \models_{u,w} A[g]$;
- $M \models_{u,v} (A = B)[g]$ iff $|A[g]|_u = |B[g]|_u$;
- $M \models_{u,v} \forall p A[g]$ iff for all $a \in U_u$, $M \models_{u,v} A[g(a/p)]$;
- $M \models_{u,v} \dagger A$ iff $M \models_{v,v} A$.

We define:

$$M \models_u A[g] \text{ iff } M \models_{u,u} A[g] \text{ (i.e., iff } u \in |A[g]|_u).$$

We say that the formula A is *valid* in the model M (in symbols, $M \models A$), if for every $u \in W$ and every assignment g into M , $M \models_u A[g]$. A is *satisfied* by the model M if for some $u \in W$ and some assignment g into M , $M \models_u A[g]$.

ACKNOWLEDGMENTS. This essay is based on the presentation I gave at the conference “David Kaplan’s Contribution to Philosophy,” San Marino, May 10–13, 2001. I am grateful to the organizers, Paolo Leonardi, Francesco Famà, and Alberto Voltolini, for inviting me. I also wish to thank David Kaplan and the other participants for encouraging remarks, as well as Wlodek Rabinowicz for comments on the penultimate

version. The essay has previously been published in A. Rojszczak, J. Cachro, and G. Kurczewski, G., eds., *Philosophical Dimensions of Logic and Science*, Synthese Library vol. 320 (Dordrecht: Kluwer, 2003). I am grateful to Springer Science and Business Media for the permission to reprint it here.

REFERENCES

- Anderson, C. A. 1980. "Some New Axioms for the Logic of Sense and Denotation: Alternative (0)." *Noûs* 14, 217–234.
- Carnap, R. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press. Second edition with supplements, 1956.
- Church, A. 1951. "A Formulation of the Logic of Sense and Denotation." In *Structure, Method, and Meaning: Essays in Honor of H. M. Sheffer*, ed. P. Henle, 3–34. New York: Liberal Arts Press.
- Church, A. 1974. "Outline of a Revised Formulation of the Logic of Sense and Denotation (Part II)." *Noûs* 8, 135–156.
- Church, A. 1976. "Comparison of Russell's Resolution of the Semantical Antinomies and That of Tarski." *Journal of Symbolic Logic* 41(4), 747–760.
- Cresswell, M. J. 1980. "Quotational Theories of Propositional Attitudes." *Journal of Philosophical Logic* 9, 17–40. Reprinted in M. J. Cresswell, *Semantical Essays: Possible Worlds and Their Rivals* (Dordrecht: Kluwer, 1988).
- Davies, M. 1981. *Meaning, Quantification, Necessity: Themes in Philosophical Logic*. London: Routledge and Kegan Paul.
- Gallin, D. 1975. *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland.
- Jubien, M., 1988. "Problems with Possible Worlds." In *Philosophical Analysis*, ed. David F. Austin, 299–322. Dordrecht: Kluwer.
- Kaplan, D. 1964. "Foundations of Intensional Logic." Dissertation, University of California, Los Angeles. University Microfilms International, Ann Arbor.
- Kaplan, D. 1975. "How to Russell a Frege-Church." *Journal of Philosophy* 72, 716–729.
- Kaplan, D. 1994. "A Problem in Possible-World Semantics." In *Modality, Morality and Belief: Essays in Honor of Ruth Barcan Marcus*, ed. W. Sinnott-Armstrong et al., 41–52. New York: Cambridge University Press.
- Kaplan, D., and R. Montague. 1960. "A Paradox Regained." *Notre Dame Journal of Formal Logic* 1, 79–90. Reprinted in Montague, *Formal Philosophy*, 1974.
- Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.
- Montague, R. 1963. "Syntactical Treatment of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability." In *Proceedings of a Colloquium on Modal and Many-Valued Logics*, Helsinki, August 23–26, 1962. *Acta Philosophica Fennica*, fasc. 16, 153–167. Reprinted in Montague, *Formal Philosophy*, 1974.
- Montague, R. 1970. "Universal Grammar." *Theoria* 36, 373–398. Reprinted in Montague, *Formal Philosophy*, 1974.

- Montague, R. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. Edited and with an introduction by Richmond H. Thomason. New Haven, Conn.: Yale University Press.
- Myhill, J. 1958. "Problems Arising in the Formalization of Intensional Logic." *Logique et Analyse* 1, 78–83.
- Myhill, J. 1979. "A Refutation of an Unjustified Attack on the Axiom of Reducibility." In *Bertrand Russell Memorial Volume*, ed. G. W. Roberts, 81–90. London: George Allen and Unwin.
- Parsons, C. 1982. "Intensional Logic in Extensional Language." *Journal of Symbolic Logic* 47, no. 2, 289–328.
- Prawitz, D. 1965. *Natural Deduction: A Proof-Theoretical Study*. Acta Universitatis Stockholmiensis, Stockholm Studies in Philosophy 3. Stockholm: Almqvist and Wiksell.
- Quine, W. V. 1953. "Three Grades of Modal Involvement." *Proceedings of the XIth International Congress of Philosophy*, vol. 14, 65–81. Amsterdam: North-Holland. Reprinted in W. V. Quine, *The Ways of Paradox and Other Essays*. Cambridge, Mass.: Harvard University Press, 1966.
- Russell, B. 1908. "Mathematical Logic as Based on the Theory of Types." *American Journal of Mathematics* 30, 222–262.
- Segerberg, K. 1973. "Two-Dimensional Modal Logic." *Journal of Philosophical Logic* 2, 77–96.
- Whitehead, A. N., and Russell, B. 1910–1913. *Principia Mathematica*. 3 vols. Cambridge: Cambridge University Press. Second edition, Cambridge: Cambridge University Press, 1925–1927.

CHAPTER 8

The Role of Variables

Kit Fine

It is generally supposed—by logicians and philosophers alike—that we now possess a perfectly good understanding of how variables work in the symbolism of logic and mathematics. Given Frege's account of their syntactic role and Tarski's account of their semantic role, it would appear that nothing more of any significance need be said. It is my belief that this common view is mistaken. There are deep problems concerning the role of variables that have never been properly recognized, let alone solved, and once we attempt to solve them, we see that they have profound implications not only for our understanding of variables but also for our understanding of other forms of expression and for the general nature of semantics.

It is my aim in this essay to explain what these problems are, to explain how they are to be solved, and to indicate the implications for the rest of semantics. I begin with an antimony concerning the role of variables which I believe any satisfactory account of our understanding of them should solve (section 1). I then argue that the three main semantical schemes currently on the market—the Tarskian, the instantial, and the algebraic—are unsuccessful in solving the puzzle (sections 2 and 3) or in providing a satisfactory semantics for first-order logic (sections 4 and 5). Finally, I offer an alternative scheme that it is capable of solving the antimony (section 6) and of providing a more satisfactory semantics for first-order logic (section 7). It is based upon a new approach to representational semantics, which I call 'semantic relationism', and I conclude by considering the implications of this approach for other topics in linguistics and the philosophy of language (section 8).

1. THE ANTINOMY OF THE VARIABLE

At the root of the present considerations is a certain puzzle. In order to state this puzzle, we shall need to appeal to the notion of *semantic role*. I do not mean this as a technical notion of a kind that one might find in a

formal semantics but as a nontechnical notion whose application may already be taken to be implicit in our intuitive understanding of a given language or symbolism. For in any meaningful expression, there is something conventional about the expression (having to do with the actual symbols or words used) and something nonconventional (having to do with the linguistic function of those symbols or words). And 'semantic role' is just my term for this linguistic or nonconventional aspect of a meaningful expression.

In fact, there will be no need, in stating the puzzle, to appeal to semantic role *per se*. For our interest will solely be in whether two expressions have the *same* or a *different* semantic role. And, of course, their having the same, or different, semantic roles is compatible with different views as to what that semantic role might be. Indeed, our puzzle would appear to arise under any reasonable conception of semantic role—be it Russellian, Fregean, or of some other kind.

Now for the puzzle. Suppose that we have two variables, say 'x' and 'y'; and suppose that they range over the same domain of individuals, say the domain of all reals. Then it appears as if we wish to say contradictory things about their semantic role. For when we consider their semantic role in two distinct expressions—such as ' $x > 0$ ' and ' $y > 0$ ', we wish to say that their semantic role is the same. Indeed, this would appear to be as clear a case as one could hope to have of a mere 'conventional' or 'notational' difference; the difference is merely in the choice of the symbol and not in its linguistic function. On the one hand, when we consider the semantic role of the variables in the same expression—such as ' $x > y$ '—then it seems equally clear that their semantic role is different. Indeed, it is essential to the linguistic function of the expression as a whole that it contains two distinct variables, not two occurrences of the same variable, and presumably this is because the roles of the distinct variables are not the same.

Generalizing from our example, we arrive at the following two claims:

Semantic Sameness (SS): Any two variables (ranging over a given domain of objects) have the same semantic role; and

Semantic Difference (SD): Any two variables (ranging over a given domain of objects) have a different semantic role.

Yet we cannot univocally maintain both (at least given that there are at least two variables!).

We might call this puzzle the 'antimony of the variable'. It was first stated by Russell, though in ontological rather than semantical fashion. He writes ([1903], sec. 93) that 'x is, in some sense, the object denoted by *any term*; yet this can hardly be strictly maintained, for different variables may occur in a proposition, yet the object denoted by *any term*, one would suppose, is unique.' It also bears a close affinity with Frege's puzzle concerning names (Frege [52]), though with variables taking the place of names.

Clearly, any resolution of the conflict must begin by locating an ambiguity in the phrase 'same semantic role'. Whatever the sense in which the variables x and y are semantically the same in ' $x > 0$ ' and ' $y > 0$ ' cannot be the sense in which they fail to be semantically the same in ' $x > y$ '. Now as a first stab toward locating the ambiguity, we might appeal to a notion of context. The variable ' x ' in the context of the formula ' $x > 0$ ' plays the same semantic role as the variable ' y ' in the context of the formula ' $y > 0$ '. On the other hand, the two variables x and y will play different semantic roles within the context of the single formula ' $x > y$ '. Thus (SD) will hold for sameness of semantic role in the *cross-contextual* sense while (SS) will hold for difference of semantic role in the *intracontextual* sense; and, given the ambiguity in the respective senses of 'same' and 'different' role, contradiction is avoided.

Natural as this response may be, it does not really solve the puzzle but merely pushes its solution back a step. For why do we say that the variables x and y have a different semantic role in ' $x > y$ '? Clearly, it has to do with the fact that the occurrence of y cannot be replaced by x without a change in semantic role; the role of x , y in ' $x < y$ ' is different from the role of x , x in ' $x < x$ '. In other words, the intracontextual difference in semantic role between the variables x and y within the single formula ' $x > y$ ' amounts to a cross-contextual difference in semantic role between the pair of variables x , y in ' $x > y$ ' and the pair of variables x , x in ' $x > x$ '. And, in general, to say that there is an intracontextual difference between x and y , in the intended sense, is just to say that there is cross-contextual difference between the pair of variables x , y and the pair x , x .

We may therefore state (SS) and (SD) in the following forms:

(SS)' there is no cross-contextual difference in semantic role between the variables x and y ; and

(SD)' there is a cross-contextual difference in semantic role between the pair of variables x , y and the pair x , x ,

using a univocal notion of semantic role throughout.

In contrast to the earlier formulation, there is now no explicit contradiction. But there is still a difficulty in seeing how the two claims, (SS)' and (SD)', might both be true. For how can there be a cross-contextual difference in semantic role between the pair of variables x , y and the pair x , x unless there is a cross-contextual semantic difference between the variables x and y themselves? What else could account for the difference in semantic role between the pairs x , y and x , x except a semantic difference in the individual variables x and y ? Or, to put it another way, if there is a semantic difference between x , y and x , x , then there must be a semantic difference between x and y ; and it is hard to see why this difference should be 'turned on' or made manifest only when the variables are used in the same context and not when they are used in different contexts.

The puzzle therefore remains, and any solution to the puzzle should either explain how (SS)' and (SD)' might be compatible, notwithstanding appearances to the contrary, or it must explain how one of (SS)' or (SD)' might reasonably be rejected.

2. THE TARSKIAN APPROACH

It might be thought that the solution to our puzzle should be sought in the various semantics that have been developed for the language of predicate logic. After all, it is presumably the aim of these semantics to account for the semantic role of the expressions with which they deal, and so we should expect them to account, in particular, for the semantic role of variables.

However, when we turn to the various semantics that have in fact been developed, we find them entirely unsuited to the purpose of solving the puzzle. Let us begin with the most familiar of these semantics, that of Tarski [36]. The reader will recall that the Tarski semantics proceeds by defining a relation of satisfaction between assignments and formulas. To fix ideas, let us suppose that the variables of our language are $x_1, x_2, \dots$ and that the domain of discourse is D . We may then take an *assignment* to be a function taking each variable of the language into an individual from D ; and the semantics will specify—by means either of a definition or of a set of axioms—what it is for each kind of formula to be satisfied by an assignment. It will state, for example, that an assignment satisfies the formula $\sim B$ just in case it fails to satisfy B or that an assignment satisfies the formula $\forall xB$ just in case every x -variant of the assignment (differing at most on the value assigned to x) satisfies B .

Now what account, within the framework of the Tarski semantics, can be given of the semantic role of the variables? There would appear to be only two options. The first is to take the semantic role of a variable to be given by its range of values (the domain D in the case above). Indeed, quite apart from the connection with the Tarski semantics, this is the usual way of indicating how a variable is to be interpreted: One specifies its range of values; and that is that.

This approach does indeed account for the fact that the semantic role of any two variables x and y (with an identical range of values) is the same. But it does nothing to account for the semantic difference between the pairs of variables x, y and x, x ; nor is any reason given for denying that there is an intuitive difference in semantic role.

The other option is to take the semantic role of a variable to be what one might call its 'semantic value' under the given semantics. The semantic values are those entities that are assigned (or that might be taken to be assigned) to the meaningful expressions of the language and with respect to which the semantics for the language is compositional. When

we examine the Tarski semantics, we see that the semantic value of an open formula (one containing free variables) might be taken to be the function that takes each assignment into the 'truth-value' of the formula under that assignment (the TRUE if the formula is satisfied by the assignment, and the FALSE otherwise) and, similarly, the semantic value of an open term might be taken to be the function that takes each assignment into the denotation of the term under that assignment. Thus the semantic value of the formula $x > 0$ (under the natural interpretation of the language) would be the function that takes any assignment into TRUE if the number it assigns to x is positive and into FALSE otherwise; and the semantic value of the term $x + y$ would be the function that takes an assignment into the sum of the numbers that it assigns to x and y . We then easily see that the Tarski semantics is compositional with respect to the semantic values as so conceived; it "computes" the semantic value of a complex expression on the basis of the simpler expressions from which it is derived.

Under this conception of semantic value, the semantic value of a variable x will be a special case of the semantic value of a term; it will be a function that takes each assignment into the individual that it assigns to x . It is therefore clear, if we identify semantic roles and semantic values, that x and y will differ in their semantic roles; for if we take any assignment that assigns different individuals to x and y (this requires, of course, that there be at least two individuals in the domain), then the semantic value of x will deliver the one individual in application to that assignment, whereas the semantic value of y will deliver the other individual in application to the assignment.

We therefore secure the semantic difference between the pairs x, y and x, x under this account of semantic role. But we are at a loss to account for the fact that the semantic role of the variables x and y is the same; nor is any reason given for disputing the intuitive identity in semantic role.

There is another problem with this approach. For although it may secure a difference between x and y (and hence between the pairs x, y and x, x), it does nothing to explain in what the difference consists. For in the last analysis, the imputed difference between the semantic values for x and y simply turns on the difference between the variables x and y themselves. The semantic value of the variable x , for example, delivers one individual rather than another in application to a given assignment simply because it is the variable that it is. Thus what we secure, strictly speaking, is not a *semantic* difference, one lying on the nonconventional side of language, but a *typographic* difference, one lying purely on the conventional side of language; and so we have done nothing to account for the *semantic* difference between x and y (or between x, y and x, x).

3. THE DENIAL OF SEMANTIC ROLE

In stating the antimony of the variable, we have presupposed that variables *have* a semantic role; and it might be thought that this is the root cause of our difficulties. For it might be thought that our understanding of variables is inseparable from their role in quantification or other forms of variable binding and that any attempt to explain the role of *free* variables, apart from their connection with the apparatus of binding, must therefore fail.

There is a familiar approach to the semantics of predicate logic that appears to lend some support to this point of view. For in attempting to provide a semantics for quantified sentences, we face a problem that is in some ways analogous to our antimony. We wish to assign a semantic value to a quantified sentence, such as $\exists x(x > 0)$, and we naturally do this on the basis of the semantic value assigned to the open sentence $x > 0$ that is governed by the quantifier. But, given that $x > 0$ and $y > 0$ are mere notational variants, they should be assigned the same semantic value, and so we should assign the same semantic value to $\exists x(x > 0)$ and $\exists x(y > 0)$ —which, clearly, is undesirable.

One solution to this problem is to deny that the semantic value of $\exists x(x > 0)$ is to be assigned on the basis of the semantic value assigned to $x > 0$, and once this line is adopted, then consistency demands that we never appeal to the semantic value of an open expression in determining the semantic value of a closed expression. The semantics for closed expressions should be ‘autonomous’ in the sense of never making a detour through the semantics of open expressions.

There are two main ways in which autonomy of this sort might be achieved. We might call them the ‘instantial’ and the ‘algebraic’ approaches. According to the first, the semantic value of a quantified sentence such as $\exists x(x > 0)$ is made to depend upon the semantic value of a closed instance $c > 0$ (e.g., $3 > 0$). The intuitive idea is that, given an understanding of a closed instance $c > 0$, we thereby understand what it is for an arbitrary individual to satisfy the condition of being greater than 0 and are thereby in a position to understand what it is for some individual or other to satisfy this condition.¹

According to the second approach, the semantic value of a quantified sentence such as $\exists x(x > 0)$ is made, in the first place, to depend upon the semantic value of the corresponding λ -term $\lambda x(x > 0)$, denoting the property of being greater than 0. Of course, this merely pushes the problem back a step, since we now need to account for the semantic value of the λ -terms. But this may be done, not in terms of the corresponding closed instances or in terms of the corresponding open formulas, but by supposing that we understand more complex ‘bindings’ in

1. Cf. Dummett [73], pp. 15–16.

terms of simpler bindings. The semantic value of $\lambda x \sim (x > 0)$, for example, will be taken to be the 'negation' of the semantic value of $\lambda x (x > 0)$, while the semantic value of $\lambda xy (x \leq y \vee y \leq x)$ will be taken to be the 'disjunction' of the semantic values of $\lambda xy (x \leq y)$ and $\lambda xy (y \leq x)$. In this way, binding may be traded in for the application of various 'algebraic' operations to the properties or relations signified by the primitive predicates of the symbolism.²

In discussing these proposals, it is important to distinguish between two different questions. The first is whether they are plausible or even viable. Can a semantics of the proposed sort be given, and, if it can, then is it faithful to the way we actually understand the symbolism? The second question is whether open expressions should even be taken to have a semantic role (which it might then be part of the aim of semantics to capture).

Of course, if there is an autonomous semantics for closed expressions, then that deprives us of one reason for thinking that open expressions have a semantic role, since they are not required to have a semantic role in order to account for the semantics of closed expressions; and I suspect that many philosophers who have been attracted to the idea of an autonomous semantics for closed expressions have been inclined, on this basis, to reject a semantic role for open expressions. However, it seems to me that there are strong independent reasons for thinking that open expressions do indeed have a semantic role.

The intuitive evidence for this appears to be overwhelming. Surely, we are inclined to think, it is at least part of the semantic role of an open term to represent a range of values. It is part of the semantic role of the term ' $2n$ ', for example, to represent any even number and part of the semantic role of the term ' $ai + b$ ' to represent any complex number. Just as it is characteristic of a closed term such as ' 2.3 ' to represent a particular individual, so it is characteristic of an open term, such as ' $2n$ ', to represent a range of different individuals; and just as the representation of a particular individual is a semantic relationship, so is the representation of a range of individuals. We would therefore appear to have as much reason to regard the representation of a range of individuals as a part of the semantic role of an open term as we have to regard the representation of a particular individual as part of the semantic role of a closed term.

But the opponent of semantic roles for open expressions may not be impressed with this evidence. For he may claim that insofar as we think of an open expression as having a semantic role, it is because we think of its variables as being implicitly bound. There is perhaps no need to think of them as being bound by a *quantifier*; thus there is no need to think of us as always using these expressions to express thoughts, but there must at least be some variable-binding operator in the background by means of which the supposed semantic role of the open expression is to be under-

2. Bealer ([79], [82], [83]) is the leading contemporary advocate of this approach.

stood. So, for example, the term ' $2n$ ' might be understood, insofar as it is seen to have a semantic role, as doing duty for the set-term ' $\{2n: n \text{ a natural number}\}$ ' or for the λ -term ' $\lambda n.2n$ ' (denoting the function from each number to its double).

I do not regard this account of the alleged semantic role of open expressions as at all plausible. The alleged semantic role of open expressions, in terms of its representing a range of values, would appear to be perfectly intelligible quite apart from the possible connection with variable-binding. Indeed, some philosophers have supposed that certain types of variable—or variable-like expression—might not even be subject to quantification or other forms of variable-binding. The schematic letters 'A' and 'B' of Quine ([52], §1.5), for example, are meant to 'stand in' for the sentences of some language and yet are not bindable on pain of supposing that the sentences are names for some special kind of entity. These schematic letters, as much as regular variables, are subject to the antimony and yet would appear to have an independent semantic role. Of course, Quine might be mistaken in his reasons for thinking that schematic letters are not bindable, but he is surely not mistaken in thinking that their semantic role can be understood apart from the connection with quantification or other forms of variable binding.

The opponent of semantic roles for open expressions also faces the awkward issue of saying what the implicit binding should be taken to be. One wants to say that a term ' $2n$ ' indifferently represents all natural numbers. Our opponent says that the term can only be regarded as having a semantic role insofar as it is implicitly bound. But by what? Two obvious candidates, when no sentential context is at hand, are the set-term ' $\{2n: n \text{ a natural number}\}$ ' and the λ -term ' $\lambda n.2n$ '; a less obvious candidate is the *arbitrary even number* of Fine ([84], [85]). Thus it must be supposed that an implicit reference and an implicit ontological commitment is made to the set or function or arbitrary object, or perhaps something else altogether. But this is both arbitrary and gratuitous. For there is no reason to suppose that the implicit reference is to one of these entities as opposed to the other, and it appears entirely irrelevant to our use of the term that it should carry any such implicit reference or commitment.

A perhaps even more decisive objection to the position arises from the consideration of semantic relationships. Not only do open expressions appear to have semantic roles, they also appear to enter into semantic relationships. For example, the 'value' of the term ' $n + 1$ ' is always greater than the 'value' of ' n ', though not of ' m '. But how is our opponent to account for these apparent semantic relationships? If he takes the variables of each term, taken on its own, to be implicitly bound, he is sunk: For then ' $n + 1$ ' will signify the successor function, say, and ' n ' the identity function, and so we will lose the special semantic relationship that holds between ' $n + 1$ ' and ' n ' as opposed to ' $n + 1$ ' and ' m '. He must therefore take the variables of the two terms to be somehow

simultaneously bound. He must say something like: What accounts for the apparent semantic relationship between ' $n + 1$ ' and ' n ' is the fact that the quantified sentence ' $\forall n(n + 1 > n)$ ' is true. But it seems bizarre to suppose that one must create this artificial context in which both terms occur in order to explain the semantic relationship between them. What kind of strange semantic relationship between the terms is it that can only be explained by embedding them within a richer language? Indeed, the proposed explanation of the semantic relationship presupposes that the relevant semantic features of the terms are preserved when they are embedded in the context of a single sentence; and so unless we had some independent way of explaining what that semantic relationship was, we would have no way to say what the presupposition was or whether it was correct.

If we are right, then the independent semantic role of open expressions is not to be denied and the antimony is not to be solved by denying that they have such a role.

4. THE INSTANTIAL APPROACH

I have so far left open the question of whether there might be an autonomous semantics for closed expressions, one not taking a detour through open expressions. But it seems to me that no such semantics is viable—or, at least, plausible. That this is so is not important for solving the antimony, but it is important for understanding the significance of a solution. For if a semantics for quantification or for other forms of variable binding must proceed by providing a semantics for open expressions, then it should provide, in particular, a semantical account of free variables, and no such account can be considered satisfactory unless it shows how the antimony is to be solved. Thus the failure of existing semantics to solve the antimony is a mark of their inadequacy; they do not properly, or fully, reveal our understanding of variables.

As I have mentioned, there are two main forms of autonomous semantics, the instancial and the algebraic, though the reasons for thinking them unsatisfactory are somewhat different. Let us begin with the instancial approach, which I regard as a closer approximation to how a satisfactory semantics should proceed.

According to this approach, it will be recalled, a closed quantified sentence such as $\exists xB(x)$ is to be understood on the basis of one of its instances $B(c)$ —the intuitive idea being that from an understanding of $B(c)$, we may acquire an understanding of what it is for an arbitrary individual to satisfy the condition denoted by $B(\)$ and that, from this, we may then acquire an understanding of what it is for this condition to be satisfied by some individual or other. But although the intuitive idea behind the proposal may be clear, it is far from clear how the proposal is to be made precise.

A certain semantic value is to be assigned to a closed instance $B(c)$ of the existential sentence $\exists xB(x)$. Let us call it a 'proposition', though without any commitment to what it is. A certain 'condition' is then to be determined on the basis of this proposition. But how? We took it to be the condition denoted by the scheme $B(\)$, which results from removing all displayed occurrences of the term c from $B(c)$. This suggests that the condition should likewise be taken to be the result of removing all corresponding occurrences of the individual denoted by c from the given proposition; indeed, we are given no other indication of how the condition might be determined. It must therefore be presupposed that there is an operation of 'abstraction' that, in application to any proposition and any occurrences of an individual in that proposition, will result in a certain condition or propositional 'form' in which the given occurrences of the individual have been removed. Once given such a form, we may then take quantified sentence $\exists xB(x)$ to predicate 'existence' of it.

Now a great deal more needs to be said about the operation of abstraction before we have a precisely formulated semantics.³ But one thing is clear. The use of such an operation in formulating the semantics for predicate logic is not compatible with an extensional approach, one in which we take cognizance only of the individuals denoted by the closed terms, the extensions of the predicates, and the truth-values of the sentences. For if there is to be a meaningful operation of abstraction, then the propositions to which it is to be applied must to some extent share in the structure of the sentences by which they are expressed; they must contain individuals in a way analogous to the way in which the sentences contain terms; and it must make sense to remove the individuals from the propositions in a way that is analogous to the removal of terms from a sentence. But clearly, there is nothing in an extensional approach that would enable us to make sense of such ideas.

So much the worse, one might think, for the extensional approach. But however sympathetic one might be to alternative semantic approaches, it is hard to believe that our current problems lie in the adoption of an extensional approach. After all, the extensionalist credentials of variables are as good as they get: They simply range over a given domain of individuals without the intervention of different senses for different individuals and without the need for different senses by which the domain might be picked out for different variables. It is therefore hard to see why the addition of variables to a language that was otherwise in conformity with extensional principles should give rise to any essential difficulties. If the extensional project fails, it cannot be because the variables carry some hidden intensional freight.

I might also note, as a historical aside, that the views of Frege [52] on the sense/reference distinction require that it be possible to provide an

3. Some hints at how a semantics of this sort might be developed are given in Fine [89], pp. 237–238. I hope to deal with the matter more fully elsewhere.

extensional semantics for the language of predicate logic, since he wanted to be able to provide a compositional semantics at the level of both sense and reference. The instantial approach is sometimes attributed to Frege (e.g., by Dummett [73], pp. 15–16), but it is only intelligible, at best, at the level of sense. And, for Frege, it is not even a viable option at the level of sense, since his views require that there be a *parallel* semantics at the level of reference.

There is in any case another, more subtle, difficulty with the instantial approach that not even the intensional form of semantics is able to solve. For it is perfectly possible to understand a quantified sentence $\exists xB(x)$ without understanding the name of any particular individual. But if the understanding of the quantified sentence is meant to be derived from an understanding of a particular instance $B(c)$, then this should not be possible.⁴ Of course, what we really wanted to say was that the understanding of $\exists xB(x)$ should be derived from our understanding of an *arbitrary* instance. We have seen that this should not be taken to mean that our understanding of $\exists xB(x)$ derives from our understanding of some particular instance, though it does not matter which. But if not this, then what?

The only reasonable view seems to be that our understanding of $\exists xB(x)$ should be taken to derive from our general understanding of $B(x)$, that is, from our understanding of the proposition expressed by $B(x)$ for any given value of the variable x . But the idea of a closed instance then falls by the wayside, and we are left with the idea of understanding the quantified sentence $\exists xB(x)$ in terms of the corresponding open sentence $B(x)$. Thus we see that, once properly understood, the instantial approach does not even provide an autonomous form of semantics.

5. THE ALGEBRAIC APPROACH

Under the alternative version of the autonomous approach, the apparatus of binding is traded in for an algebra of operations, with the apparatus serving, in effect, as a device to indicate how the semantic value for a whole sentence is to be generated from the properties and relations expressed by the primitive predicates that it contains. In this case, there is no difficulty in making the semantics precise or in presenting it in extensional form. But there is a difficulty in seeing how to extend it beyond the standard symbolism of predicate logic.

One difficulty of this sort arises from the use of quantifiers that apply to several variables at once, though in no set order. We might have a quantifier ‘always’, for example, that implicitly quantifies over all vari-

4. In laying out the instantial approach, I have assumed that the domain of quantification may be restricted and that the instantial term c should denote an individual from the domain. But the present objection does not rest on either of these assumptions.

ables ‘in sight’. Let us symbolize it by ‘ $\forall$ ’ (without attached variables) and take ‘ $\forall A$ ’ to indicate that ‘ A ’ holds no matter what values are assumed by the free variables occurring in ‘ A ’. The question now arises as to how the proponent of the algebraic approach is to understand a sentence such as ‘ $\forall (x + y = y + x)$ ’.⁵ Clearly, he must understand it in terms of the application of a universality operator to a λ -term constructed from ‘ $x + y = y + x$ ’. But which λ -term? There would appear to be only two options:

- (i) It is a λ -term, ‘ $\lambda xy(x + y = y + x)$ ’ or ‘ $\lambda yx(x + y = y + x)$ ’, in which the variables attached to the λ -symbol are taken to occur in a set order;
- (ii) it is a λ -term, ‘ $\lambda(x + y = y + x)$ ’, in which the λ -symbol is taken, like the quantifier-symbol $\forall$, to apply to all free variables in sight, though in no set order.

But the first option gives a more specific meaning to the sentence than it actually appears to possess. We must therefore either arbitrarily adopt one interpretation over another or attribute to the sentence an indeterminacy in meaning that it does not seem to have. And do not think that the relevant order might be given by the standard alphabetic order of the variables or by the order in which they occur in the formula that follows the quantifier, for we might take the variables to be symbols—such as ‘ a ’ and ‘ λ ’—which are not alphabetized in any given order and the formula might be written in a nonlinear notation from which no set order of occurrence can be discerned.⁶

The second option is clearly more faithful to our semantic intentions, but it plays havoc with the idea of distributing the binders across the logical connectives. For, on pain of reintroducing an arbitrary order on the variables, a λ -term such as $\lambda(x > y)$ will have to symbolize a relation that is neutral between the ‘biased’ relations symbolized by $\lambda xy(x > y)$ and $\lambda yx(x > y)$.⁷ But this means that, when we push the binder λ through $\lambda(x > y \vee y > x)$ in order to obtain the disjunction of the relations symbolized by $\lambda(x > y)$ and $\lambda(y > x)$, we will lose track of the alignment between the variables in the two disjuncts and will therefore be unable to distinguish, in the way we should, between $\lambda(x > y \vee y > x)$ and $\lambda(x > y \vee x > y)$.

A perhaps more serious difficulty of the same sort arises from the use of modal and other intensional operators. Suppose that we add an operator ‘ $\Box$ ’ for necessity to the symbolism for first-order logic and that we take the interpretation of the quantifiers to be ‘actualist’—ranging, in each possible world, over the objects that exist in that world. Consider

5. Mathematicians sometimes use the strict equality sign ‘ $\equiv$ ’ in this way: ‘ $x + y \equiv y + x$ ’, for example, is used to indicate the universal truth of ‘ $x + y = y + x$ ’. I might note that a similar difficulty arises for branching quantifiers, since the different ‘branches’ of the quantifier do not occur in a set order.

6. As with the notation for fractions or determinants: $\frac{x}{y} = \frac{x}{y}$ and $\begin{vmatrix} x & y \\ u & v \end{vmatrix} = \begin{vmatrix} x & y \\ u & v \end{vmatrix}$.

7. Some accounts of what a neutral relation might be are considered in Fine [00].

now the algebraic treatment of $\Box(x = x)$ and of $\Box\exists y(y = x)$. In conformity with the algebraic approach, $\lambda x(x = x)$ should be understood to signify the result of applying some operation, call it 'necessitation', to the property (of self-identity) signified by $\lambda x(x = x)$, and $\lambda x\Box\exists y(y = x)$ should be understood to signify the result of applying this same operation to the property (of existence) signified by $\lambda x\exists y(y = x)$. But under an actualist interpretation of the quantifier (and hence also of λ -binding), $\lambda x(x = x)$ and $\lambda x\exists y(y = x)$ will be modally indistinguishable—they will be true, in each possible world, of the individuals that exist in that world. And this is plausibly taken to imply, under any natural understanding of $\Box$ as a modal operator, that $\Box(x = x)$ and $\Box\exists y(y = x)$ should also be modally indistinguishable—which, of course, they are not, since the first is satisfied by any individual whatever while the second is satisfied only by individuals that necessarily exist.

There are further difficulties for the approach, even in regard to the unadorned form of first-order logic; for it requires us, at almost every turn, to make arbitrary decisions about the interpretation of the symbolism that have no counterpart in our actual understanding of the symbolism. Let me merely give one illustration. When we push the λ -operators through logically complex formulas, we will eventually reach λ -terms of the form $\lambda x_1 x_2 \dots x_m F y_1 y_2 \dots y_n$, in which the binder $\lambda x_1 x_2 \dots x_m$ governs an atomic formula $F y_1 y_2 \dots y_n$ (and where some of the variables $x_1, x_2, \dots, x_m$ may, of course, be the same as one another and the same as some of the variables $y_1, y_2, \dots, y_n$). How, then, are these terms to be interpreted?

Presumably, in keeping with the algebraic approach, we should take $\lambda x_1 x_2 \dots x_m F y_1 y_2 \dots y_n$ to signify the result of applying some operation to the relation F signified by the predicate F . So, for example, $\lambda x F x x$ will indicate the reflexive version of F , while $\lambda y x F x y$ will signify the converse of F (at least, if $\lambda x y F x y$ signifies F itself). But what, in general, is this operation? It must presumably be determined on the basis of the relative disposition of the variables $x_1 x_2 \dots x_m$ in the binder and of the variables $y_1 y_2 \dots y_n$ in the atomic formula. But there are different ways in which it might be so determined, and nothing to choose between them. Consider $\lambda x z F x x z$, for example. We could take this to be the result of first forming the 'converse' $\lambda x y z F x y z$ of $\lambda x y z F x y z$ and then forming the reflexive version $\lambda x z F x x z$ of the converse, or we could take it to be the result of first forming the reflexive version $\lambda x z F x x z$ of $\lambda x y z F x y z$ and then forming the converse $\lambda x z F x x z$ of the reflexive version; and similarly, and to a much more complicated extent, for other cases.

These choices do not correspond to anything in our actual understanding of the symbolism, and so we face the awkward choice of making the interpretation either arbitrarily specific or unacceptably indeterminate. For this reason, the algebraic approach is best viewed as an attempt to see the symbolism of first-order logic as something which it is not. What it provides, in effect, is a translation from a language with variables to one

without variables. A λ -term such as $\lambda x x F x x z$, for example, may be taken to be equivalent in meaning to a term $\text{Refl}(\text{Conv}(F))$, indicating the application of reflexive and converse operations to the relation signified by F , and all appeal to variables in the target language is thereby made to disappear. But in making the transition to a variable-free notation, not only are we forced to make arbitrary decisions about how the translation should go; we thereby lose what is most distinctive about the use of variables. For instead of being treated as devices of reference, albeit of a special sort, they are treated as more or less oblique ways to indicate the application of various operations within a calculus of relations. The problem of understanding our use of variables is not explained, but ignored.

6. THE RELATIONAL APPROACH

I now wish to indicate how I think the antimony is to be solved and how a more satisfactory semantics for the symbolism of first-order logic might be developed on the basis of the solution.

I agree with the autonomous approach in taking the formulation of the antimony to embody a false presupposition. But the false presupposition lies not in the attribution of a semantic role to free variables but in the presumption that there is conflict between the sameness in semantic role of x and y , on the one hand, and the difference in semantic role of x , y and x , x , on the other ((SS)' and (SD)' above).

There are, I believe, two things that stand in the way of our seeing how these attributions of sameness and difference might be reconciled. The first concerns a possible ambiguity in the notion of semantic role. We have already had occasion to distinguish between sameness or difference of semantic role *across* contexts and *within* a given context. But there is, I believe, another ambiguity in the notion of semantic role that might stand in the way of seeing how the attributions of sameness and difference are to be reconciled.

This may be brought to light by considering the following argument against their reconciliation. Suppose, in conformity with (SD)', that the semantic roles of the pair x , y and of the pair x , x are not the same. Then it may be argued that the semantic roles of the individual variables x and y cannot be the same, in contradiction to (SS)'. For x , when paired with x , has the same semantic role as x , x whereas y , when paired with x , does not have the same semantic role as x , x . The variables therefore differ in respect of whether their pairing with x gives something with the same semantic role as x , x ; and this difference, presumably, is a semantic difference.

It is not to be denied that there is a semantic difference of the presumed sort between x and y . But it is not in this sense that we wish to deny that there is a semantic difference between x and y . To see how

this is so, let us distinguish between the *intrinsic* (or nonrelational) and the *extrinsic* (or relational) semantic features of an expression. The intrinsic semantic features of an expression, in contrast to its extrinsic semantic features, do not concern its semantic relationship to other expressions. Thus it will be an intrinsic semantic feature of the predicate 'doctor' that it is true of doctors but not an intrinsic semantic feature that it is synonymous with 'physician'. Likewise, the intrinsic semantic features of a *pair* of expressions will consist of those semantic relationships between the expressions that do not concern their semantic relationship to yet further expressions. Thus it will be an intrinsic semantic feature of the pair 'doctor' and 'physician' that they are synonymous, though not that they are both synonymous with 'licensed medical practitioner'.

Now what the above argument shows is that if there is an intrinsic semantic difference between the pairs of variables x, y and x, x , then there will be an extrinsic semantic difference between the individual variables x and y themselves (concerning the relationship of each to the variable x). But in asserting that the semantic role of x and y is the same, we only wish to assert that their intrinsic semantic features are the same; and in asserting that the semantic roles of x, y and x, x are different, we mean to assert that their intrinsic semantic features are different. Thus the present difficulty will not arise as long as we always take semantic role to be intrinsic.

The other impediment to achieving reconciliation rests upon a mistake in doctrine rather than upon a failure to recognize a distinction. Let it be granted that the relevant notion of semantic role is both cross-contextual and intrinsic. Still, it might be asked, how can there be a difference in the (intrinsic) semantic relationships holding between each of two pairs of expressions without there being a difference in the intrinsic semantic features of the expressions themselves? Thus given a difference in relationship between the pairs 'doctor', 'dentist' and 'doctor', 'doctor' (with the one being synonymous and the other not), there must be a difference in meaning between 'doctor' and 'dentist'. Similarly, given that there is a difference in the semantic relationships holding between the pairs x, y and x, x , there must be a difference in meaning between x and y . Indeed, without such a difference, we would have, in each case, a pair of synonymous expressions, and it is hard to see how they could give rise to a difference in relationship. But if the variables differ in meaning, then one must have one meaning and the other another meaning; and this will be a difference in their intrinsic semantic features. According to this view, then, there can be no difference in intrinsic semantic relationship without a difference in intrinsic semantic feature. All differences in meaning must be attributable to intrinsic differences, and any attempt to reconcile the attributions of semantic sameness and difference must therefore fail.

It has to be acknowledged that this view of meaning—what we might call 'semantical intrinsicism'—is extremely persuasive; and I suspect that it may be a view of this sort that has stood in the way of seeing the

semantic behavior of variables for what it is. However, once this behavior is examined, without theoretical preconception, for what it is, then it will be seen to be clearly in conflict with the view; and, since the actual behavior of the variables is not to be denied, it is the more theoretical view that should be given up.

Suppose again, to fix our ideas, that we are dealing with a language that contains the variables $x_1, x_2, x_3, \dots$. How, then, is their semantic behavior to be described? For simplicity, let us suppose that their semantic behavior is to be described entirely in extensional terms, since nothing in our argument will turn upon allowing intensional elements to appear in the description.

Now we should certainly specify the range of values each variable can assume and, of course, as long as the language is 'one-sorted', the range of values for each variable will be the same. Now normally, in providing some kind of semantical description of the variables, nothing more is said (perhaps because of the grip of intrinsicalist doctrine). But something more does need to be said. For we should specify not only which values each *single* variable can assume, when taken on its own, but also which values *several* variables can assume, when taken together. We should specify, for example, not only that x_1 can assume the number 2 as a value, say, and x_2 the number 3 but also that x_1 and x_2 can *simultaneously* assume the numbers 2 and 3 as values; and, in general, we should state that the variables take their values independently of one another, that a variable can take any value from its range regardless of which values the other variables might assume.

What is important to appreciate here is that it does not follow, simply from the specification of a range of values for each variable, which values the variables can *simultaneously* assume. One might adopt the proposal of Wittgenstein (1922), for example, and disallow distinct variables from taking the same value; or, at the other extreme, one might insist that distinct variables should always assume the same value (treating them, in effect, as strict notational variants of one another); and there are, of course, numerous other possibilities. Thus the fact that distinct variables assume values in complete independence of one another is an additional piece of information concerning their semantic behavior, one not already implicit in the specification of their range.

However, once we have specified the range and the independence in value, then we will have a complete description of the semantic behavior of the variables; there is nothing more (at least at the extensional level) to be said about their role.⁸ But if this is so, then it is clear that the intrinsicalist doctrine, *no difference in semantic relationship without a difference in semantic feature*, will fail. For the intrinsic semantic features of any two variables will be the same—they will in effect be given by the

8. At least at the extensional level and if we are only concerned with the semantic behavior of the variable-types and not also of the variable-occurrences.

specification of their range, whereas the intrinsic semantic features of the pairs x_1, x_2 , say, and x_1, x_1 will be different, since the former will assume *any* pair of values from the given range while the latter will assume only *identical* pairs of values. If we are merely informed of the intrinsic semantic features of two variables, then we cannot tell whether they assume their values independently of one another (should they be distinct) or whether they always assume the same value (should they be the same).

It is thus by giving up the intrinsicalist doctrine, plausible as it initially appears to be, that the antimony is to be solved. We must allow that any two variables will be semantically the same, even though pairs of identical and of distinct variables are semantically different, and we should, in general, be open to the possibility that the meaning of the expressions of a language is to be given in terms of their semantic relationships to one another. Formally, the situation is analogous to failures in the identity of indiscernibles. Consider, for example, the distinct but indiscernible spheres of Max Black [70]. Just as there is no intrinsic spatial difference between the two spheres (once we reject absolute space), so there is no intrinsic semantic difference between two variables. And just as there may be an intrinsic spatial difference between two pairs of spheres—since identical spheres will be coincident whereas distinct spheres will not be—so there may be an intrinsic semantical difference between pairs of variables. From this perspective, then, the lesson to be drawn from the antimony is that semantics provides another, though less familiar, example of an aspect of ‘reality’ in which things can only be distinguished in terms of their relations to one another, but not in terms of their intrinsic features.

7. RELATIONAL SEMANTICS FOR FIRST-ORDER LOGIC

I now wish to consider how the relational solution to the antimony is capable of yielding a more satisfactory semantics for the symbolism of first-order logic. I deal first with the case of free variables, which is somewhat more straightforward, and then consider the complications that arise from the use of bound variables.

I take it to be the aim of a semantics for a given language to account for the semantic behavior of its expressions. The way this is usually done is by assigning a semantic value to each meaningful expression of the language, with the semantic value of a complex expression being determined on the basis of the semantic values of the simpler expressions from which it is syntactically derived. The semantic value of each expression is naturally taken to correspond to its intrinsic semantic role; and so, given the truth of semantic intrinsicalism, the assignment of semantic values to the expressions of a language should then be sufficient to determine their semantic behavior.

But we have seen that the doctrine of semantic intrinsicism should be abandoned; there are intrinsic semantic relationships between expressions that are not grounded in their intrinsic semantic role. This means that the aim of semantics needs to be reconsidered. For it is no longer sufficient to assign semantic values to expressions; we should also take account of semantic relationships between expressions that may not be grounded in their intrinsic semantic features. Let us use 'semantic connection' as the formal counterpart, within a semantics, to the informal notion of semantic relationship (just as 'semantic value' is the formal counterpart, within a semantics, to the informal notion of semantic role). A semantics should concern itself, then, with the assignment of semantic connections, and not merely with the assignment of semantic values.

Indeed, we should go further still. It is not that semantic inquiry, as it is usually conceived, needs to be *supplemented* by an account of semantic connection. For given that there are irreducible semantic relationships among expressions, one can reasonably expect that the semantic value of certain complex expressions will depend on the semantic connections among the simpler expressions from which it is derived, and not merely on their semantic values. Thus semantic connection should *replace* semantic value as the principal object of semantic inquiry, and the aim of semantics should be to determine the semantic connections among expressions on the basis of the semantic connections among simpler expressions.

Following through this strategy, the semantics for a given language will eventually terminate in the 'lexical semantics', which accounts for the behavior of those expressions, the 'lexical items', that are not syntactically derived from any other expressions. However, the lexical semantics, like the semantics as whole, must now be taken to assign semantic connections to the lexical items, and not merely semantic values. It is on the basis of these 'primitive' semantic connections that the semantic connections among all expressions of the language will be determined.

If we are to apply this general approach to the language of first-order logic, then we must first have some conception of what we want the semantic connections to be. To this end, we should associate with any sequence of meaningful expressions of a first-order language a unique semantic connection (which we treat as the sum total, as it were, of the various semantic relationships that may hold among those expressions). Suppose, in the first place, that we are given a sequence of closed expressions—say, ' $2 + 3$ ', ' $2 > 3$ ', ' 6 '. Then we take the semantic *value* (not connection) of the sequence to be the corresponding sequence of semantic values of the members of the sequence—in the case at hand (and under the obvious interpretation), this will be the sequence 5, FALSE, 6. Suppose now that we are given a sequence of open expressions—say, ' $x + y$ ', ' $x > y$ ', ' z '. Then under different assignments of values to the variables, such a sequence will, in an obvious way, take different semantic values. Under the assignment of 2 to ' x ', 3 to ' y ', and 6

to 'z', it will take the sequence 5, FALSE, 6 as its semantic value, for example, while under the assignment of 4 to 'x', 2 to 'y', and 2 to 'z', it will take the sequence 6, TRUE, 2 as its semantic value. We now take the *semantic connection* on a sequence of expressions (either open or closed) to be the set of semantic values that the sequence of expressions is capable of assuming under different assignments of individuals to its variables. For example, the semantic connection on 'x + y', 'x > y', 'z' will include the sequences 5, FALSE, 6 and 6, TRUE, 2. Thus a semantic connection on a sequence of open expressions will indicate how the semantic values of those expressions are capable of varying with respect to one another.

Given the current conception of semantic connection, the aim of the semantics should be to determine the semantic connection on any given sequence of expressions—starting with the lexical semantics, for the very simplest expressions, and then successively working through more and more complicated forms of expression. The lexical semantics is, for the most part, straightforward: extensions should be assigned to predicates, denotations to constants, and functions to function symbols. However, we now include variables within the lexicon, and so the lexical semantics should also specify the semantic connection on any sequence of variables. Suppose that we are given the sequence of variables x, y, x, y, for example. Then in conformity with our understanding that distinct variables take values independently of one another and that identical variables take the same value, the semantic connection on this sequence should be the set of all quadruples a, b, c, d of individuals from the domain for which $a = c$ and $b = d$. And, in general, the semantic connection on the variables $x_1, x_2, \dots, x_n$ should be taken to be the set of all n -tuples $a_1, a_2, \dots, a_n$ of individuals from the domain for which $a_i = a_j$ whenever $x_i = x_j$ ($1 \leq i < j \leq n$). It is exactly at this point that the relational aspect of the semantics first becomes manifest.⁹

We now need rules for extending the semantic connections to more complicated sequences of expressions. Consider, by way of example, the sequence 'x + y', 'x', 'y'. We would like to determine the semantic connection of this sequence on the basis of the semantic value of the operator '+' and the semantic connection on the sequence 'x', 'y', 'x', 'y'. So suppose that the semantic value of '+' is + (for addition) and that we are given the semantic connection on 'x', 'y', 'x', 'y' (consisting of all quadruples a, b, c, d of individuals from the domain for which $a = c$ and $b = d$). Then the semantic connection on the sequence 'x + y', 'x', 'y' should be taken to consist of all triples of the form $a + b, c, d$, where a, b, c, d is a member of the semantic connection on 'x', 'y', 'x', 'y'. Since the semantic connection on 'x', 'y', 'x', 'y' will consist of quadruples of the form a, b, a, b , the semantic connection on 'x + y', 'x', 'y' will consist of

9. The lexical semantics should also let us combine the connection on two sequences of expressions—such as x, y and $+, 0$ —which are independent of one another.

triples of the form $a + b$, a, b ; and so the required coordination between ' $x + y$ ', ' x ', and ' y ' will be achieved by means of the preexisting coordination in value between the variables in ' x ', ' y ', ' x ', ' y '.

More generally, let us suppose that we have a complex term $ft_1t_2 \dots t_n$ and that we wish to determine the semantic connection on a sequence of expressions involving the term. The general form of such a sequence will be $D_1, D_2, \dots, D_p, ft_1t_2 \dots t_n, E_1, E_2, \dots, E_q$ (with $p, q \geq 0$). Now, under the standard approach to semantics, the semantic value of the complex term $ft_1t_2 \dots t_n$ will be determined on the basis of the semantic value f of its function symbol f and the semantic values $a_1, a_2, \dots, a_n$ (under a given assignment) of its argument-terms $t_1, t_2, \dots, t_n$; and this semantic value will be taken to be the result $f(a_1, a_2, \dots, a_n)$ of applying the function f to the arguments $a_1, a_2, \dots, a_n$. This suggests that the semantic connection on the sequence $D_1, D_2, \dots, D_p, ft_1t_2 \dots t_n, E_1, E_2, \dots, E_q$ should be determined on the basis of the semantic connection on $D_1, D_2, \dots, D_p, f, t_1, t_2, \dots, t_n, E_1, E_2, \dots, E_q$ (with $ft_1t_2 \dots t_n$ giving way, as before, to $f, t_1, t_2, \dots, t_n$); and $d_1, d_2, \dots, d_p, b, e_1, e_2, \dots, e_q$ will belong to the semantic connection on $D_1, D_2, \dots, D_p, ft_1t_2 \dots t_n, E_1, E_2, \dots, E_q$, that is, $ft_1t_2 \dots t_n$ will be capable of taking the value b when $D_1, D_2, \dots, D_p, E_1, E_2, \dots, E_q$ take the values $d_1, d_2, \dots, d_p, e_1, e_2, \dots, e_q$ just in case, for some individuals $a_1, a_2, \dots, a_n$ for which $b = f(a_1, a_2, \dots, a_n)$, $t_1, t_2, \dots, t_n$ are capable of taking the values $a_1, a_2, \dots, a_n$ when $D_1, D_2, \dots, D_p, E_1, E_2, \dots, E_q$ take the values $d_1, d_2, \dots, d_p, e_1, e_2, \dots, e_q$, that is, just in case, for some individuals $a_1, a_2, \dots, a_n$, for which the n -tuple $d_1, d_2, \dots, d_p, f, a_1, a_2, \dots, a_n, e_1, e_2, \dots, e_q$ belongs to the semantic connection on $D_1, D_2, \dots, D_p, f, t_1, t_2, \dots, t_n, E_1, E_2, \dots, E_q$.

The above rule is easily extended to the case of atomic formulas, and a similar rule may be given in the case of truth-functionally complex formulas. The semantic connection on $\sim A, E$, for example, will consist of all those pairs π', e for which π, e is a member of the semantic connection on A, E and π' is the 'opposite' truth-value to π . Note how the present semantics distinguishes between the semantic connections on ' $x > y$ ' and on ' $x > x$ '. The first should consist of the truth-values TRUE and FALSE, while the second should contain only the truth-value FALSE. The semantic connections on ' x ' and on ' y ' are the same. But the semantic connections on x, y and on x, x are different (one consisting of all pairs a, b of individuals from the domain and the other consisting of all identical pairs a, a of individuals from the domain); and it is upon this difference in connection on the *pairs* of variables that the semantic distinction between the atomic formulas is made to turn.

Let us now turn to the case of the quantifiers, which raise additional complications. The obvious way of evaluating a sequence containing a quantified formula—such as $\exists xA(x), E$ —is in terms of the connection on $x, A(x), E$. If, for a fixed semantic value of E , $A(x)$ is true for the assignment of some individual from the domain to ' x ', then $\exists xA(x)$

should be taken to be true and otherwise should be taken to be false. Thus the pair TRUE, e will belong to the connection on $\exists xA(x)$, E just in case, for some individual a from the domain, the triple a , TRUE, e belongs to the connection on x , $A(x)$, E (and similarly for the pair FALSE, e).¹⁰

But consider how such a rule applies to a sequence of the form $\exists xA(x)$, x —say to $\exists x(x > 0)$, x (where the domain of quantification is the set of all natural numbers). The pair TRUE, 0 will belong to the semantic connection on $\exists x(x > 0)$, x just in case, for some natural number n , the triple n , TRUE, 0 belongs to the semantic connection on x , $x > 0$, x . But since the first and third variables in x , $x > 0$, x are the same, the first and third components of any triple in its semantic connection will be the same. It follows that no triple of the form n , TRUE, 0 can belong to the semantic connection on x , $x > 0$, x and so the pair TRUE, 0 will not belong to the semantic connection on $\exists x(x > 0)$, x —contrary to our intentions.

The problem is that we do not want the bound occurrences of the variable x in $\exists xA(x)$, x to be ‘coordinated’ with the free occurrence. However, our method of evaluation requires that they be coordinated, since $\exists xA(x)$, x is evaluated in terms of x , $A(x)$, x . What makes the problem especially acute is that we wish to subject $\exists x(x > 0)$, x to essentially the same method of evaluation as $\exists y(y > 0)$, x , since they are mere notational variants of one another, and yet we also want to subject $\exists y(y > 0)$, x to the straightforward evaluation in terms of y , $y > 0$, x .

The way out of the impasse, I believe, is to give up the assumption that all occurrences of the same variable should be treated in the same way. We have so far assumed that different free occurrences of the same variable should always take the same individual as value; this is, indeed, a reasonable default assumption to make. However, when a free occurrence of a variable was originally bound, becoming free only in the process of evaluation, then we should no longer assume that it is coordinated with those occurrences of the same variable that were either originally free or had their origin in a different quantifier.

This means that, for the purpose of evaluating a sequence of expressions, we should explicitly indicate which of the free occurrences of a given variable are to be coordinated and which not. In the sequence x , $x > 0$, x , for example, we should distinguish between the cases in which none of the occurrences of x are to be coordinated, in which all are, or in which only two are. (The reader might picture these coordinating links as lines connecting one occurrence of the variable to another—as in the familiar ‘telegraphic’ notation for quantifier binding.)

10. I have assumed that the quantified formula $\exists xA(x)$ contains no free variables. If it does, then we can only properly evaluate the formula by considering its connection to the free variables that it contains; the above rule is in effect relativized to an assignment of individuals to those variables.

This then provides us with the means of evaluating $\exists x(x > 0)$, x in terms of x , $x > 0$, x , in strict analogy with the evaluation of $\exists y(y > 0)$, x in terms of y , $y > 0$, x . However, the variables in the evaluating sequences should be subject to the appropriate pattern of coordination; the first two occurrences of x in x , $x > 0$, x should be coordinated with one another (though not with the third occurrence), and the first two occurrences of y in y , $y > 0$, x should likewise be coordinated. In general, whenever we are evaluating a quantified formula $\exists xA(x)$ in terms of its components x , $A(x)$ within the context of a sequence, the designated occurrences of x should be taken to be coordinated with one another, though not with any other occurrences of x that may happen to be present in the sequence.

The semantic rule for the variables should also be modified, since the original formulation presupposed the default coordination on any two occurrences of the same variable. The rule should now be taken to apply to a *coordinated* sequences of variables, ones subject to some or another pattern of coordination. And the requirement for a sequence of individuals $a_1, a_2, \dots, a_n$ from the domain to be in the semantic connection of a coordinated sequence of variables $x_1, x_2, \dots, x_n$ is that a_i should be the same as a_j whenever x_i and x_j are coordinated ($1 \leq i < j \leq n$).

The other semantic rules should also be modified to take account of the fact that the sequences of expressions are now coordinated; and once this is done, the semantics, as we have stated it, is essentially complete.

The relational semantics has several advantages over the alternatives. First and foremost, it embodies a solution to the antimony: the intrinsic semantic features of x and y (as given by the semantic connections on those variables) are the same, though the intrinsic semantic features of the pairs x, y and x, x (again, as given by the semantic connections on those pairs) are different. But the semantics is also more satisfactory, in various ways, *as* a semantics. In contrast to the autonomous approaches, it assigns a semantic role to open expressions; in contrast to the instantial approach, it can be given an extensional formulation; and in contrast to the algebraic approach, it is based upon a credible direct method of evaluation. Finally, in contrast to the Tarski semantics, it is nontypographic; by going relational, we avoid having to incorporate the variables themselves (or some surrogate thereof) into the very identity of the entities that the semantics assigns to the open expressions of the language.

The reader may be troubled by the complexity of our account, especially in comparison with the Tarski semantics. But one should bear in mind that the more abstract a semantics for a given language, the more removed it is from the underlying syntax, the more complicated it is likely to be. Indeed, it seems to me that the complexity of our own account is of a desirable sort. For it appears not to be a mere artifact of the approach but to reflect a genuine complexity in our understanding. The relational approach forces us to take seriously the pattern of coordination among the variables; these patterns must be explicitly used as

syntactic inputs to the semantic method of evaluation. And it is hard to believe that their semantic significance, as revealed by the semantics, is not somehow integral to our understanding of the variables. From among the alternatives, it is only the algebraic approach that takes explicit account of the phenomenon of coordination, but it does this in a manner that is completely artificial and not in keeping with the variable's referential role.

8. IMPLICATIONS

Our approach to the semantics of variables has various broader implications which I should briefly discuss. As I have already indicated, the proper semantic treatment of variables provides a clear and convincing refutation of the doctrine of semantic intrinsicism. Since this doctrine is presupposed by the two leading approaches to the semantical study of language, the Fregean and the Russellian, one would expect semantic relationism to have broad implications for how these approaches are to be pursued once the relationist aspects of meaning are taken into account.

It is important, in this connection, to distinguish between the present version of relationism and the more familiar doctrine of semantic holism. Our own version of relationism is meant to serve as an *extension* to the usual forms of representational semantics. We take it to be a legitimate task of semantics to assign referents or senses to expressions and then consider whether semantics, as so conceived, might also require the postulation of semantic relationships. Semantic holism, on the other hand, is meant to serve as an *alternative* to the usual forms of representational semantics. The task of assigning referents or senses to expressions is abandoned in favor of the task of providing an account of their inferential or 'conceptual' role. Thus the two approaches are relational in very different ways, with the one staying squarely within the representational framework and the other locating the relational aspects of meaning outside of that framework.

Semantic relationism might not be so interesting, even as a general doctrine, if the case of variables provided the only instance of its truth. But I believe that, once one is open to the possibility of relationism, one can then see many other ways in which it might be applied. Perhaps one of the most interesting concerns the distinction between what might be called 'strict' and 'accidental' coreference. A Millian will claim that two coreferential proper names, such as 'Cicero' and 'Tully', will have the same intrinsic semantic features, since there is no more to their meaning than their referents. However, it is still possible, I believe, for the Millian to distinguish the semantic relationship between 'Cicero' and 'Cicero' (which is one of *strict* coreference) from the semantic relationship between 'Cicero' and 'Tully' (which is one of *accidental* coreference). Thus strict and accidental coreference will be semantic relationships for the Millian that are not

grounded in the intrinsic semantic features of the expressions between which they hold. Moreover, once this distinction is made, the Millian has new means of dealing with Frege's puzzle ([52]), with Kripke's puzzle concerning belief ([79]), and with other general issues concerning the relationship of semantics to cognition.

A somewhat similar distinction may be drawn between *strict* and *accidental* synonymy. Thus 'doctor' and 'doctor' will be strictly synonymous, while 'doctor' and 'physician' will be only accidentally synonymous. This distinction is one that should be accepted by the Fregean as well as the Millian and, like the parallel distinction between strict and accidental reference, provides us with new means of dealing with Mates's puzzle ([72]), Moore's paradox of analysis, and other such issues.

Relational ideas are also important to the project of constructing a detailed semantics for ordinary language. The most obvious application is to anaphora, which should be treated as a form of strict coreference (or synonymy). But there are numerous correlative terms and constructions that also appear to call for a relational treatment. Coordinated uses of 'Mr. Smith' and 'Mrs. Smith', for example, should be taken to refer to husband and wife; and in a construction such as 'the better it is for him, the worse it is for her', it is clear that the semantic value of 'the better' should somehow be coordinated with the semantic value of 'the worse'.

Finally, there are further applications to be made to first-order logic itself. For the apparatus of coordination provides us with much greater flexibility in formulating different forms of semantics for the use of variables, and it provides us, in particular, with the means to formulate a referential semantics for intensional first-order logic that is tolerant of failures of substitutivity and existential presupposition that naturally arise once one attempts to quantify into various forms of intensional context.¹¹

ACKNOWLEDGMENTS. This essay is based upon Fine [03]; and many of the topics are further discussed in Fine [07]. I should like to thank the members of seminars at Saint Andrews University, New York University, and UCLA for helpful discussion.; and I am especially grateful to Ruth Chang and Louis Derosset for their comments on an earlier version of the essay. But my greatest debt of gratitude is to David Kaplan. His work on reference has had a deep impact on my own thinking; and some of his remarks in Kaplan [90] already anticipate the relationist point of view.

REFERENCES

- Almog J., Perry J., and Wettstein H. [89] *Themes from Kaplan*. Oxford: Oxford University Press.

- Bealer G. [79] "Theories of Properties, Relations, and Propositions." *Journal of Philosophy* 76, 634–648.
- Bealer G. [82] *Quality and Concept*. Oxford: Clarendon Press.
- Bealer G. [83] "Completeness in the Theory of Properties, Relations and Propositions." *Journal of Symbolic Logic* 48(2), 415–426.
- Black M. [70] "The Identity of Indiscernibles." In *Universals and Particulars*, ed. M. Loux), 204–216. Garden City, N.Y.: Doubleday.
- Dummett M. [73] *Frege: Philosophy of Language*. London: Duckworth.
- Fine K. [84] "A Defence of Arbitrary Objects." *Aristotelian Society*, suppl. vol. 57 (1983), 55–77. Reprinted in *Varieties of Formal Semantics*, ed. F. Landman and F. Veldman. Dordrecht-Cinnaminson: Foris, 1984.
- Fine K. [85] *Reasoning with Arbitrary Objects*. Oxford: Blackwell.
- Fine K. [89] "The Problem of De Re Modality." In Almog, *Themes from Kaplan* [89], 197–272.
- Fine K. [98] "A Defense of Cantorian Abstraction." *Journal of Philosophy*, 98(12), 599–634.
- Fine K. [00] "Neutral Relations." *Philosophical Review* 109(1), 1–33.
- Fine K. [03] "The Role of Variables." *Journal of Philosophy* 100(12), 605–631.
- Fine K. [07] "Semantic Relationism." Oxford: Blackwell.
- Frege G. [52] "On Sense and Reference." Translation of "Über Sinn und Bedeutung." In *Translations from the Philosophical Writings of Gottlob Frege*, ed. P. T. Geach and M. Black. Oxford: Clarendon Press.
- Kaplan D. [90] "Words." *Proceedings of the Aristotelian Society* 64 (Suppl.), 93–119.
- Kripke S. [79] "A Puzzle about Belief." In *Meaning and Use*, ed. A. Margalit. Dordrecht: Reidel.
- Mates B. [72] "Synonymity." In *Semantics and the Philosophy of Language*, ed. L. Linsky, 111–138. Urbana: University of Illinois Press.
- Quine W. V. O. [52] *Methods of Logic*. London: Routledge and Kegan Paul.
- Russell B. [03] *Principles of Mathematics*. London: Allen and Unwin.
- Tarski A. [36] "Der Wahrheitsbegriff in den Formalisierten Sprachen." *Studia Philosophica* 1, 261–405. Originally published in 1933. Reprinted as "The Concept of Truth in Formalized Languages," in A. Tarski, *Logic, Semantics and Metamathematics*, trans. J. H. Woodger (Oxford: Clarendon Press, 1955).
- Wittgenstein L. [1922] *Tractatus-Logico-Philosophicus*. London: Routledge-Kegan.

This page intentionally left blank

Part III

LANGUAGE AND THOUGHT

This page intentionally left blank

CHAPTER 9

Reference, Inference, and the Semantics of Pejoratives

Timothy Williamson

1. INTRODUCTION

Two opposing tendencies in the philosophy of language go by the names of 'referentialism' and 'inferentialism,' respectively. In the crudest version of the contrast, the referentialist account of meaning gives center stage to the referential semantics for a language, which is then used to explain the inference rules for the language, perhaps as those which preserve truth on that semantics (since a referential semantics for a language determines the truth-conditions of its sentences). By contrast, the inferentialist account of meaning gives center stage to the inference rules for the language, which are then used to explain its referential semantics, perhaps as the semantics on which the rules preserve truth. On pain of circularity, we cannot combine both directions of explanation.

Of course, this simple opposition may not survive more refined definitions of 'referentialism' and 'inferentialism.' Not all of those willing to apply one or the other term to their views would accept the crude characterizations just given. Nevertheless, the two styles of theorizing often function as rivals in practice. There is a corresponding contrast in the philosophy of thought, concerning the reference of concepts, rather than of expressions of a public language, and inferential connections between thoughts, rather than between sentences.

If inferentialists appealed only to inference rules, they could not hope to explain how many words refer to extralinguistic objects, or how language is used in interaction with the extralinguistic environment. We may therefore assume that, in addition to intralinguistic rules such as the introduction and elimination rules for the logical constants in a Gentzen-style system of natural deduction, inferentialists may also advert to 'language-entry' rules that connect perceptual states to moves in the language game, and 'language-exit' rules that connect moves in the language game to nonlinguistic actions. Inferential roles are thereby

generalized as conceptual roles.¹ The details do not matter here, for the overall picture is the same. The rules by which speakers use the language have explanatory primacy. The referential semantics is a kind of epiphenomenon. At least on the simplest versions of inferentialism, understanding a word is equated with using it according to the appropriate rules, not with having propositional knowledge of its referential meaning: practice is prior to theory.

At first sight, inferentialism may appear to be a conservative or quietist view: since whatever rules the speech community uses thereby determine a semantics on which those very rules come out valid, the community can never rightly be criticized for using invalid rules. By contrast, referentialists appear not to be barred from assigning a semantics to the language of a community on which invalid rules are in use.

In a classic paper, Arthur Prior (1960) showed that not all inference rules are self-validating in the way that simple-minded inferentialists imagine. For consider these rules for a new binary sentence connective 'tonk':

<i>Tonk-Introduction</i>	<i>Tonk-Elimination</i>
$\frac{A}{A \text{ tonk } B}$	$\frac{A \text{ tonk } B}{B}$

Tonk-Introduction is just like one half of the standard introduction rule for disjunction; Tonk-Elimination is just like one half of the standard elimination rule for conjunction. Taken together, they allow one to infer any conclusion from any premise. Given that the 'tonk'-free part of the language contains at least one true sentence and at least one false sentence (with respect to a given context), no assignment of meaning to 'tonk' makes both Tonk-Introduction and Tonk-Elimination truth-preserving.

In an almost equally classic reply to Prior, Nuel Belnap (1962) pointed out in effect that we can find fault with the rules for 'tonk' without deviating from the inferentialist paradigm. For if we start with a 'tonk'-free language and a system of inference rules in which not every sentence can be inferred from every other sentence, and then extend the language by adding 'tonk' and the system of rules by adding Tonk-Introduction and Tonk-Elimination, the result is not a *conservative extension* of the original system. An extension is conservative if and only if whenever a conclusion in the original language follows according to the extended system of rules from premises also in the original language, that conclusion already followed according to the original system of rules from those premises.² In other words, the restriction of the

1. Although perception and action are typically extralinguistic, it does not follow that they are extraconceptual: on some views, perceptual content is already conceptual, and guiding intentions make intentional action conceptual too.

2. For the purposes of generalizing the notion of a conservative extension to systems including language-entry and language-exit rules, perceptions and actions can be treated on a par with sentences.

extended consequence relation to the original language is just the original consequence relation; the new rules do not interfere with inferential relations between old sentences. The idea of a conservative extension is evidently defined in terms available to inferentialists; it adverts to no referential semantics for the original or extended language.

Belnap suggested a further constraint on the rules with which an expression is introduced if they are to define it properly. The rules R(E) should provide a *unique characterization* of the new expression E, in the sense that if we introduce *two* new expressions E1 and E2, governed by the corresponding rules R(E1) and R(E2), respectively, then those rules should make E1 and E2 inferentially equivalent in an appropriate sense.³ Informally, failure of unique characterization is supposed to show that the rules are too weak to define the new expression adequately; failure of conservative extension is supposed to show that they are too strong to do so.

Belnap's combined constraints of conservativeness and uniqueness are closely related, although not quite equivalent, to Michael Dummett's requirement of harmony on the rules for a logical connective for them to count as self-justifying (1991: 246–251). Not all inferentialists accept Belnap's constraints. For example, Robert Brandom denies that nonconservativeness is an automatic vice in the rules for a new connective, since the new inferences between sentences of the old language might be good, useful ones (1994: 127–130; 2000: 71–72). His brand of inferentialist assesses inferential practices case by case, in a more nuanced or ad hoc way.

'Tonk' does not provide a fully satisfying example of an inferentialist critique of an inferential practice, for no community could seriously make unrestricted use of both Tonk-Introduction and Tonk-Elimination, on pain of trivializing their use of the whole language. Nor does the example provide a clear case of a defective concept, since one could plausibly deny that 'tonk' stands for a concept at all (Peacocke 1992: 21). The question therefore arises: are any naturally occurring inferential practices defective in this way? For Dummett, the ordinary use of negation according to classical logic provides just such a case. His claim is hard to assess, because highly sensitive to the form in which classical logic is presented (Rumfitt 2000). However, Dummett (1973: 397, 454) also points to a class of expressions in natural language that seem to exhibit nonconservativeness of a peculiarly simple and vicious kind: pejoratives, such as terms of ethnic abuse.

Dummett takes as his example the word 'Boche,' perhaps because its air of anachronism and its associations with anti-German propaganda in Britain and France at the outbreak of the First World War isolate us from its derogatory implications more clinically than mere quotation marks can do. It is harder to bring oneself to write down or pronounce more

3. For more discussion of the unique characterization requirement, see Williamson (1987/1988), McGee (2000), and references therein.

currently potent expressions of abuse. On Dummett's view, the meaning of the word 'Boche' is constituted by rules of inference along these lines (1973: 454):

<i>Boche-Introduction</i>	<i>Boche-Elimination</i>
$\frac{x \text{ is a German}}{x \text{ is a Boche}}$	$\frac{x \text{ is a Boche}}{x \text{ is cruel}}$

As Dummett points out, these rules induce a nonconservative extension of the 'Boche'-free language, since they permit the inference from 'He is a German' to 'He is cruel,' which presumably could not be made without them.⁴ They also exhibit lack of harmony in Dummett's sense. Thus ethnic abuse rests on inadequate proof-theory.

Dummett's account of pejoratives has been widely accepted. Although Brandom does not endorse conservativeness as a universal constraint, he accepts Dummett's description of the rules for 'Boche' (1994: 126; 2000: 69–70). Similarly, it is relied on by Paul Boghossian, although he is not in general an inferentialist outside the realm of logic (2003: 241–242).

Part of the interest of Dummett's account is that pejoratives are commonly occurring, clearly meaningful terms. As Boghossian points out, it would be highly implausible to deny that the word 'Boche' expresses a concept; surely xenophobes use sentences in which it occurs to express complete thoughts, however bad those thoughts are (2003: 242–243).⁵ One might even regard Dummett's account not merely as defusing a potentially tricky case for the inferentialist program but as providing it with a positive success, by elegantly explaining in inferentialist terms just what is wrong with pejorative expressions and concepts.

There is thus some point to investigating whether Dummett's inferentialist account of pejoratives in natural language is in fact correct. One main contention of this essay is that it is quite mistaken. An alternative account will be provided within the framework of referentialist semantics. Little will be said explicitly about the overall upshot for the inferentialist-referentialist debate, the terms of which may be too inchoate for fruitful engagement at that level of abstraction. However, an issue will tentatively be raised about the bearing of pejoratives on another vexed and programmatic debate, that over the relative priority of language and thought.

4. It is not clear that 'cruel' exactly captures the pejorative connotation of 'Boche,' but it will do for the sake of argument. Dummett has 'barbarous and more prone to cruelty than other Europeans' (1973: 454).

5. Boghossian's claim that 'Boche' expresses a genuine concept even though its rules cannot be interpreted as truth-preserving represents a change of view from that presented in Boghossian 2002.

2. UNDERSTANDING PEJORATIVES

What is required for understanding a pejorative expression or grasping a pejorative concept? On the simplest development of Dummett's account, one understands 'Boche' if and only if one is disposed to apply Boche-Introduction and Boche-Elimination. Boghossian writes, "Plausibly, a thinker possesses the concept *boche* just in case he is willing to infer according to [Dummett's rules]" (2003: 241–242).

Let us first note a grammatical point. 'Boche' can be used both as a noun and (perhaps less commonly) as an adjective. Like 'German,' it is restricted to people when used as a noun but not when used as an adjective. Although Dummett does not specify which use he has in mind, his rules have been formulated here only for the use of 'Boche' as a noun (witness the indefinite article; by formulating Dummett's rules for 'Boche' without the article, Boghossian (2003: 242) treats it as an adjective). The adjectival use clearly has no direct connection with the inference from '*x* is Boche' to '*x* is cruel,' for not even the wildest xenophobe will infer from 'That camera is Boche,' 'That camera is cruel.' Thus Dummett's account of 'Boche' is incomplete, since it does not cover the adjectival use. That does not show that it is incorrect about the use of 'Boche' as a noun.

Is a disposition to reason according to Boche-Introduction and Boche-Elimination necessary and sufficient for understanding the noun 'Boche'? I trust that I am one counterexample and that you are another. Unlike someone who thinks that the word 'Boche' means *box*, we fully understand the word, for we understand sentences that xenophobes utter in which it occurs. We know what 'Boche' means. We find racist and xenophobic abuse offensive because we understand it, not because we fail to do so. Yet we are unwilling to infer according to both Boche-Introduction and Boche-Elimination. Similarly, imagine a reformed xenophobe who once was willing to infer according to those rules but now has seen the error of his ways, while vividly recalling with shame what it was like to shout xenophobic abuse. He still remembers what 'Boche' means, and so understands the word.

People who were never brought up or inclined to feel anti-German sentiment do not even have an underlying disposition to infer according to those rules which they inhibit as a result of training in political correctness, respect for proof-theoretic hygiene, or a general patina of civilization. They have no such disposition whatsoever, even though they can imagine what it is like to have such a disposition. Nevertheless, they understand 'Boche'; they do not merely have the ability to imagine what it is like to understand it. They need not even have the conditional disposition to infer according to the rules if they used 'Boche' at all: rather, their disposition might be to use 'Boche' in ways subversive of the rules if they used it at all.

Since understanding the word 'Boche' (with its present meaning) is presumably sufficient (although not necessary) for having the concept that 'Boche' expresses, it follows that a willingness or disposition to reason according to Dummett's rules is equally unnecessary for having that concept.

The objection does not depend on the details of Dummett's rules. Consider any set of rules that an inferentialist proposes for 'Boche.' If they are logically unobjectionable—more specifically, if they constitute a conservative extension of a civilized 'Boche'-free system of rules—then the inferentialist has no account of what is objectionable about 'Boche.' For the primary objection to the use of the word does not depend on how it is applied: even the statement 'If Lessing was a Boche then he was a Boche' is objectionable. On the other hand, if the rules are logically objectionable—if they constitute a nonconservative extension of the civilized 'Boche'-free system of rules—then millions of civilized people are counterexamples to the claim that a willingness or disposition to infer according to the rules is necessary for understanding 'Boche,' or for having the concept that it expresses. Either way, the account is inadequate. Can a civilized person say directly *what* 'Boche' means, what concept it expresses? If one says "'Boche" means *Boche*' or "'Boche" expresses the concept *Boche*,' one is in danger of using the word 'Boche' in its italicized occurrence, not just mentioning it, and thereby of committing oneself to the xenophobic abuse. But if one cannot say (even homophonically) what the word means, does one really understand it? However, that 'Boche' is used in its ordinary sense in the italicized occurrences is far from clear. The italics may be tantamount to sense quotation, a device by which one refers to the ordinary meaning of the expression, not to its ordinary referent (Kaplan 1969). That would not be a straightforward, committing use of the pejorative. In any case, a difficulty in saying (even homophonically) what a word means need not imply a difficulty in understanding it. Consider a tribe in which men and women are supposed to use quite different words for some things. A woman might happen to know the men's word for something but not know the women's word for it; she might not even have a synonym in women's vocabulary. She understands the men's word perfectly well but cannot say what it means, or what concept it expresses, without using that very word herself, and thereby committing a linguistically inappropriate act, breaking the rules. Similarly, non-xenophobes can understand the word 'Boche' quite well, whatever proprieties interfere with making their understanding explicit.

The inferentialist may retreat to the claim that to understand a term is to know what the rules for it are, not to be disposed to reason according to those rules. One knows the rules of the 'Boche' game but is unwilling, perhaps for that very reason, to play it. On this conception, understanding a term consists not so much in using it appropriately as in knowing how to use it appropriately (by its rules). Although the inferentialist may

insist that this knowing how that one has never been inclined to put into practice is still practical knowledge, it is unclear what is behind this slogan. One applies one's knowledge of what pejoratives mean in understanding xenophobic discourse: if one is applying knowledge how to infer by the 'Boche' rules, despite not inferring by those rules, it needs to be explained in what way that knowledge is much more practical than the semantic knowledge that the referentialist invokes.⁶

A further problem faces even the watered-down inferentialist account of understanding pejoratives. Someone might grow up in a narrow-minded community with only pejorative words for some things, in particular with only the pejorative 'Boche' for Germans. He might understand 'Boche' as other xenophobes do without understanding 'German' or any equivalent nonpejorative term. He would be unacquainted with Boche-Introduction and any similar rule. Thus not even knowing how to infer according to Boche-Introduction is necessary for understanding 'Boche,' or for having the concept that it expresses.

3. THE REFERENCE OF PEJORATIVES

For inferentialists, the rules for an expression determine whatever is determined about its reference (given a context of utterance). Suppose that the referential semantics for a language has already been determined. Now a new expression *E* is added to the original language, subject to inferential rules *R*(*E*). The natural inferentialist account of the reference of *E* is this. If the hypothetical assignment of *X* as the reference of *E* makes *R*(*E*) truth-preserving, and no other assignment does, then *E* does indeed refer to *X* (for classical inferentialists, the standard introduction and elimination rules for the usual truth-functors uniquely determine their referential semantics, in other words their truth-tables, in this way). If several different hypothetical assignments of reference to *E* all make *R*(*E*) truth-preserving, then *E* is indeterminate in reference between all those assignments (traditional supervaluationist semantics might be used to combine them all into a single semantic upshot with truth-value gaps). But if no hypothetical assignment of reference to *E* makes *R*(*E*) truth-preserving, then *E* does not refer. The simplest sentences in which *E* occurs are therefore either uniformly false or uniformly truth-valueless, although it may still be allowed that *E* has a sense, and expresses a concept.

Since there are noncruel Germans, no assignment of reference to 'Boche' makes both Boche-Introduction and Boche-Elimination truth-preserving. Therefore, on the picture just sketched, 'Boche' does not

6. Brandom (1994: 25–26) proposes to explain knowledge-how in terms of knowledge-that. For an argument that knowledge-how is just a special case of knowledge-that, see Stanley and Williamson (2001).

refer. Thus atomic ascriptions of 'Boche' are either uniformly false or uniformly truth-valueless. Either way, 'The Kaiser was a Boche' and 'Kitchener was a Boche' have the same semantic status.⁷ Presumably, if 'He is a Boche' is always false, then 'He is not a Boche' is always true (when the negation is not Horn's metalinguistic negation [1989]), while if 'He is a Boche' is always truth-valueless, then 'He is not a Boche' is always truth-valueless too.

A complete account of the reference of a term should apply to its occurrence in modal contexts. To extract such implications from Boche-Introduction and Boche-Elimination, one would have to treat them as applicable in modal contexts, and therefore as implying that *necessarily* all Germans are cruel. It is doubtful that many xenophobes go that far, but for the sake of simplicity this further dubious aspect of Dummett's account will be ignored. For present purposes, we can identify the reference of a general term with its extension in the actual world (rather than, for example, with a property).

Surprisingly, on the inferentialist account of reference determination above, even consistent xenophobes who think that all Germans are cruel should conclude that Dummett's rules for 'Boche' fail to determine its reference uniquely. On their view, the set of all Germans is a subset of the set of all cruel people, but not vice versa (they think that Germans are far from the only cruel foreigners), so many sets have the set of all Germans as a subset and are in turn subsets of the set of all cruel people.⁸ If any such intermediate set is assigned to 'Boche' as its extension, both Boche-Introduction and Boche-Elimination are truth-preserving. In particular, for the xenophobe, both rules are truth-preserving if the extension of 'Boche' is the set of all Germans, but are equally truth-preserving if the extension of 'Boche' is the set of all cruel people. In Belnap's terms, Boche-Introduction and Boche-Elimination together are not uniquely characterizing. On this view, the logically sophisticated xenophobe (the British or French counterpart of Frege?) should conclude that 'Boche' is indeterminate in reference between all these different sets. But that result badly misrepresents the xenophobe's conception. For the xenophobe thinks that the term 'Boche' is literally false of non-Germans, however cruel they are. Dummett's account fails to explain why xenophobes say things like 'He's as cruel as a Boche, but he isn't a Boche—he's a Bulgar.' Dummett gives no more weight to one of his rules than to the other, but the xenophobic use of 'Boche' does not treat them symmetrically.

7. The change from present to past tense is immaterial.

8. A xenophobe might apply the term 'Boche' to someone he knew not to be German whom he regarded as displaying a distinctively German kind of cruelty, but that is a recognizably metaphorical use. The xenophobe would not apply 'Boche' even metaphorically to someone he knew not to be German whom he regarded as displaying a distinctively non-German kind of cruelty.

We can begin to compare the status of the two rules by considering each of them separately. Any introduction rule determines an elimination rule in harmony with it: the elimination rule specifies consequences of applying the target expression that match the conditions for applying it that the introduction rule specifies. In the simplest case, such as Boche-Introduction, the introduction rule has a single premise and a single conclusion; the corresponding elimination rule has the conclusion of the introduction rule as its single premise and the premise of the introduction rule as its single conclusion; the two rules are mutual converses. Similarly, any elimination rule determines an introduction rule with which it is in harmony: the introduction rule specifies conditions for applying the target expression that match the consequences of applying it that the elimination rule specifies. In the simplest case, such as Boche-Elimination, the elimination rule has a single premise and a single conclusion; the corresponding introduction rule has the conclusion of the elimination rule as its single premise and the premise of the elimination rule as its single conclusion; again, the two rules are mutual converses. If the introduction and elimination rules for an expression are not in harmony, then either the elimination rule gets more out of applying the expression than the introduction rule put in or the introduction rule puts more into applying the expression than the elimination rule gets out. But if the rules are in harmony, then the sufficient condition for applying the expression that the introduction rule supplies is equivalent to the necessary condition for applying it that the elimination rule supplies; the rules satisfy Belnap's constraints of conservativeness and uniqueness, and uniquely determine the reference of the expression (given a referential semantics for the rest of the language). Let us consider Dummett's two rules in turn.

(A) Boche-Introduction determines this elimination rule in harmony with it:

*Boche-Elimination**

x is a Boche
 x is a German

Boche-Elimination* merely allows one to infer from ' x is a Boche' what Boche-Introduction allows one to infer ' x is a Boche' from, and Boche-Introduction merely allows one to infer ' x is a Boche' from what Boche-Elimination* allows one to infer from ' x is a Boche.' Consequently, if the two rules are added to a theory in a language without 'Boche,' they yield a conservative extension: anything in the original language provable with them was already provable without them. Boche-Elimination* is the elimination rule that naturally corresponds to Boche-Introduction. Together, Boche-Introduction and Boche-Elimination* determine the set of all Germans as the extension of 'Boche': it is the only candidate extension that makes both rules truth-preserving (for every value of the variable).

This is the assignment of reference that naturally corresponds to Boche-Introduction. The set of all Germans is the smallest set that makes Boche-Introduction truth-preserving. On this interpretation, 'Lessing was a Boche' is true, because Lessing was a German, but 'Nero was a Boche' is false, because Nero was not a German. Boche-Elimination is invalidated: every noncruel German yields a counterexample.

(B) Boche-Elimination determines this introduction rule with which it is in harmony:

*Boche-Introduction**

x is cruel
 x is a Boche

(If the variable can take nonpersons as values, then 'cruel' in both rules should be expanded to 'a cruel person.') Boche-Introduction* merely allows one to infer ' x is a Boche' from what Boche-Elimination allows one to infer from ' x is a Boche,' and Boche-Elimination merely allows one to infer from ' x is a Boche' what Boche-Introduction* allows one to infer ' x is a Boche' from. Consequently, if the two rules are added to a theory in a language without 'Boche,' they yield a conservative extension: anything in the original language provable with them was already provable without them. Boche-Introduction* is the introduction rule that naturally corresponds to Boche-Elimination. Together, Boche-Introduction* and Boche-Elimination determine the set of all cruel people as the extension of 'Boche': it is the only candidate extension for 'Boche' that makes both rules truth-preserving (for every value of the variable). This is the assignment of reference that naturally corresponds to Boche-Elimination. The set of all cruel people is the largest set that makes Boche-Elimination truth-preserving. On this interpretation, 'Lessing was a Boche' is false, because Lessing was not a cruel person, but 'Nero was a Boche' is true, because Nero was a cruel person. Boche-Introduction is invalidated: every noncruel German yields a counterexample.

On the picture of Boche-Introduction and Boche-Elimination as carrying equal weight in determining reference, or the lack of it, for 'Boche,' (A) and (B) have exactly the same status; the advantages and disadvantages of the two treatments of 'Boche' are symmetrically related. But that result is wildly implausible. Intuitively, Boche-Elimination* is just as intimately linked to the meaning of 'Boche' as Boche-Introduction is. By contrast, Boche-Introduction* has no standing at all; even most xenophobes would reject it, for they think that Germans are not the only cruel foreigners. Again, although both 'Lessing was a Boche' and 'Nero was a Boche' are regrettable utterances, the former seems to have a far stronger claim to literal truth than the latter has. The xenophobic abuse is preserved in the negations of those sentences, but 'Lessing was not a Boche' seems to have a far weaker claim to literal truth than 'Nero was not a Boche' has. Thus Boche-Introduction trumps Boche-Elimination.

A further asymmetry between the two rules is observable in practice. Xenophobes typically treat Boche-Elimination as defeasible because they allow that there are a few good Germans ('Some of my best friends are Boche'), whereas they treat both Boche-Introduction and Boche-Elimination* as indefeasible. Such asymmetries undermine inferentialist attempts to give explanatory primacy to Dummett's account of 'Boche.'

The natural if tentative conclusion to draw from the preceding considerations is that, far from suffering reference failure or massive indeterminacy of reference, 'Boche' has the same reference as 'German.' All the dictionaries that I have consulted define 'Boche' to that effect. With minor variations of no present concern, they all define 'Boche' as 'German,' while glossing the word as 'contemptuous' or 'derogatory.' On the lexicographers' view, 'Boche' applies to all Germans, however kind, and to no non-Germans, however cruel. Thus Boche-Introduction and Boche-Elimination* are exceptionlessly truth-preserving, while Boche-Elimination is very far from truth-preserving. Accounts that make Boche-Introduction and Boche-Elimination the basic rules of use for 'Boche' are therefore highly implausible.

The inferentialist attempt to derive the reference of pejoratives from Dummett's account leads to results inconsistent with the consensus of descriptive linguists not involved in the philosophical dispute. Could the inferentialist improve on Dummett's account by proposing a different set of inference rules? The general difficulty is this. If the inference rules for 'Boche' constitute a conservative extension of a civilized system of rules for the 'Boche'-free part of the language, then they do not explain what is offensive about even such statements as 'If Lessing was a Boche, then he was a Boche.' On the other hand, if the rules for 'Boche' constitute a nonconservative extension of the civilized system, then the inferentialist account of reference determination seems to imply that ' x is a Boche' is not true for any value of ' x ,' which is empirically implausible. One strategy for the inferentialist is to say that the indefeasible rules for 'Boche' are Boche-Introduction and Boche-Elimination*, while Boche-Elimination is a merely defeasible rule. On the natural hypothesis that indefeasible rules trump defeasible ones in determining reference, the extension of 'Boche' would be the set of all Germans, but 'Boche' would still be offensive in virtue of the defeasible inference (by the indefeasible Boche-Introduction and the defeasible Boche-Elimination) from ' x is a German' to ' x is cruel.' This is not totally dissimilar from the view developed below. But it still distorts the picture. For, as already noted, the word 'Boche' might be used as a pejorative for Germans in a narrow-minded community with no nonpejorative equivalent of 'German.' That would not affect its referential status. Thus inferences involving 'German' are not crucial to determining the reference of 'Boche.'

Can some inferentialist story utterly different from Dummett's be told about pejoratives? At least, anything like the account of 'Boche' on which

inferentialists have actually relied badly misrepresents the meaning of pejoratives. A fresh start is needed.

4. AN ALTERNATIVE ACCOUNT

In Fregean terminology, the lexicographers' account of 'Boche' is in effect that it has the same reference as 'German' but a different tone. Indeed, since the differences between 'Boche' and 'German' apparently play no role in determining reference, and so make no difference to the way in which the terms contribute to the truth-conditions of sentences in which they occur, a Fregean might even count 'Boche' and 'German' as having the same sense. Frege himself gives just such an account of another pejorative term (1979: 140): 'cur' has the same sense and reference as 'dog' but a different tone. According to Frege, 'This dog howled the whole night' and 'This cur howled the whole night' express the same thought (the same sense). They differ in that only the latter conveys an attitude of contempt for the dog on the part of the speaker. Frege denies that 'This cur howled the whole night' expresses the further thought that the speaker has such an attitude, because the absence of the attitude would not falsify the sentence.⁹

It might be objected to Frege's account that if the words 'dog' and 'cur' have the same sense, then, on Frege's own account of propositional attitude ascriptions, the sentences 'Mary believes that every dog is a dog' and 'Mary believes that every dog is a cur' must have the same truth-value, however much Mary loves dogs. Mary, a fully competent speaker of English, assents to 'Every dog is a dog'; will she assent to 'Every dog is a cur'? If she agrees that 'Every dog is a cur' is true but misleading, we can surely agree that 'Mary believes that every dog is a cur' is also true but misleading. What if Mary, under the influence of too much inferentialist reading, claims that 'Every dog is a cur' lacks a truth-value in virtue of its badly matched introduction and elimination rules? If she believes that every dog is a cur, why does she not assent to 'Every dog is a cur,' for surely she knows that it expresses her belief that every dog is a cur? Such problems undermine, if anything, Frege's simple account of propositional attitude ascriptions, not his claim that pairs like 'cur' and 'dog' have the same truth-conditional meaning. For similar problems arise even for pairs of synonyms with the same tone. Kripke (1979) gives the example of the synonymous natural kind terms 'furze' and 'gorse.' A speaker might learn the two words on different occasions, from ostension of examples, and acquire normal competence with both without being sure that they refer to exactly the same kind of plant. Thus

9. The example involves complications about the proper treatment of indexicals and of descriptive elements in demonstratives, but Frege's general line is clear. Dummett (1973: 84–89) gives a nuanced account of Frege on tone.

'He believes that all furze is furze' and 'He believes that all furze is gorse' appear to differ in truth-value. It does not follow that 'furze' and 'gorse' are not synonyms after all. Such problems show nothing special about pejoratives. Whatever the right account of propositional attitude ascriptions, it is compatible with the Fregean view that 'cur' and 'dog' differ in tone but not sense (truth-conditional meaning).

Unfortunately, Frege's category of tone is too miscellaneous to take us very far in the analysis of the example. It includes stylistic differences, such as that between 'sweat' and 'perspire,' which are significantly unlike the difference between 'Boche' and 'German.' Frege discusses his examples by speaking unhelpfully of the images and feelings that the words evoke in hearers. Nevertheless, the classification of the difference between 'Boche' and 'German' as truth-conditionally irrelevant is at least a useful start, for it respects intuitive distinctions over which the original inferentialist account rides roughshod.

What needs explaining is this. Competent English speakers know, or are in a position to know, that 'German' and 'Boche' have the same reference, and therefore that 'Lessing was a German' and 'Lessing was a Boche' have the same truth-value. If educated, such speakers know, or are in a position to know, that both sentences are true. Nevertheless, although such speakers are willing to assert 'Lessing was a German,' they are not willing to assert 'Lessing was a Boche,' even on reflection, unless they are xenophobes. I know that 'Lessing was a Boche' is true, but I refuse to assert 'Lessing was a Boche.' Why?

The natural answer is that to assert 'Lessing was a Boche' would be to *imply* that Germans are cruel, and I do not want to imply that, because the implication is both false and abusive. Since the false implication that Germans are cruel does not falsify 'Lessing was a Boche,' it is not a logical consequence of 'Lessing was a Boche.' Rather, in Grice's terminology, 'Lessing was a Boche' has the *conventional implicature* that Germans are cruel, in much the same way that 'Helen is polite but honest' has the conventional implicature that there is a contrast between Helen's being polite and her being honest.¹⁰ Just as 'Lessing was a Boche' and 'Lessing was a German' differ in conventional implicatures while being truth-conditionally equivalent, so too 'Helen is polite but honest' and 'Helen

10. See Grice (1989: 41) and (1961: sec. 3). The example of the truth-conditionally irrelevant difference between 'and' and 'but' goes back to Frege (1879: §7), but he misdescribes the difference by requiring the contrast to consist in the unexpectedness of what follows 'but' (Dummett (1973: 86). Frege (1979: 140) gives this analogy for someone who uses the word 'cur' without feeling the contempt that it implies: 'If a commander conceals his weakness from the enemy by making his troops keep changing their uniforms, he is not telling a lie; for he is not expressing any thoughts, although his actions are calculated to induce thoughts in others.' This captures the conventional nature of the implicature only if the troops flout conventions as to which uniforms they are permitted to wear. A question by Owen Greenhall about the relation between the pairs 'Boche'/'German' and 'but'/'and' in a class at Edinburgh first interested me in the present line of thought about pejoratives.

is polite and honest' differ in conventional implicatures while being truth-conditionally equivalent. In Grice's terms, conventional implicatures are *detachable*, because they can differ between truth-conditionally equivalent sentences (in the same context). But they are not easily *cancellable*, for someone who says 'Lessing was Boche, although I do not mean to imply that Germans are cruel' merely adds hypocrisy to xenophobia; equally deviant would be an utterance of 'Helen is polite but honest, although I do not mean to imply that there is any contrast between her being polite and her being honest.' By contrast, Gricean conversational implicatures are easily cancelable but not detachable. Whether one says 'She is either in Paris or Rome' or something truth-conditionally equivalent to it, the maxims of conversation generate the implicature that one does not know which city she is in; nevertheless, one can easily cancel it by adding, 'I know which city she is in, but I am not going to tell you.'¹¹

The implicature that 'Boche' carries must be conventional rather than purely conversational, for if there were not already a significant difference between the words 'Boche' and 'German,' the norms of conversation would not generate any difference in implicature between 'Lessing was a Boche' and 'Lessing was a German.' Of course, there can be nonsemantic sociological differences between terms with the same reference. For instance, the expressions *E* and *E*^{*} both refer to *X*, but *E* predominates in the dialect of a social group *G* whose members tend to view *X* positively, while *E*^{*} predominates in the dialect of a social group *G*^{*} whose members tend to view *X* negatively. But it does not follow that a member of *G* who uses *E* thereby conversationally implies (perhaps by manner) something positive about *X*, or that a member of *G*^{*} who uses *E*^{*} thereby conversationally implies (perhaps by manner) something negative about *X*. For *E* may simply be the default, neutral term for *X* in *G*, smoothly available even to the few members of *G* who view *X* negatively, while *E*^{*} is the default, neutral term for *X* in *G*^{*}, smoothly available even to the few members of *G*^{*} who view *X* positively. In any case, such a drastic sociological difference is empirically implausible for 'Boche' and 'German.' Those who used 'Boche' were not presenting themselves as members of a social group in which anti-German feeling was commonly known to predominate; they were insulting Germans much more directly. The failure of cancelability for 'Boche' confirms this difference. One does not cancel the implicature by saying 'Lessing was a Boche, but I'm not one of those German-hating people who use "Boche".'

11. Perhaps logically equivalent sentences in the same context can differ in conversational implicature: compare 'She said that she was there' and 'She said that she was there, and either she was lying or she wasn't.' Those two sentences still differ in truth-conditions by a very fine-grained standard (intensional isomorphism). Even by a more fine-grained standard of truth-conditional equivalence, detachability may be only a rough guide to conversational implicature. By choosing long words one might conversationally imply that the children must not be told.

The conventional implicatures of 'Boche' and 'but' are preserved under embedding in more complex sentences. For example, 'If Lessing was a Boche then he was a European' still conventionally implicates that Germans are cruel, and 'If Helen is polite but honest then she is honest' still conventionally implicates that there is a contrast between her being polite and her being honest. The implicatures are present just as strongly in nonindicative sentences, such as 'Is he a Boche?' 'Give this to a Boche!' 'Is she polite but honest?' And 'Be polite but honest!' In this respect, such conventional implicatures are like presuppositions rather than logical consequences. However, the relevant sense of 'presupposition' is closer to Stalnaker's than to Strawson's: although a presupposition modifies the context of utterance, its failure does not deprive the sentence of a truth-value.¹² In the case of 'Boche,' the very use of the word generates the xenophobic implicature, irrespective of its position in the sentence.¹³

The false implicature of 'Lessing was a Boche' is not merely that Lessing was cruel. For otherwise the corresponding implicature of 'Himmler was a Boche' would merely be that Himmler was cruel, which is commonly known to be true. Then I could decently assert 'Himmler was a Boche': but I cannot. Thus even a singular ascription of 'Boche' carries the false general implicature that Germans are cruel. But since xenophobes treat Boche-Elimination as a defeasible rule, the implicature is not that all Germans without exception are cruel. The implicature is expressed by 'Germans are cruel' read as a generic sentence, meaning something like 'There is a tendency for Germans to be cruel.' A xenophobe may easily say 'He's a Boche, but he's not cruel—he's one of the few decent ones.'¹⁴

The implicature that 'Boche' carries is not merely about the speaker's psychological state. In particular, what is implicated is not merely that the speaker *believes* that Germans are cruel; such an implicature might

12. Contrast Stalnaker (1999: 38–40, 47–62) with Strawson (1952: 175–179). For a related view of conventional implicature, see Karttunen and Peters (1979).

13. If the mere use of the pejorative word is what generates the implicature, then the compositional properties of this kind of conventional implicature may contrast with those of the usual paradigms of presupposition. For example, a conditional does not automatically inherit the presuppositions of its consequent: whereas 'He has stopped beating his wife' presupposes 'He once beat his wife,' the conditional 'If he once beat his wife then he has stopped beating his wife' lacks that presupposition (for a related approach to compositional features of conventional implicature, see Karttunen and Peters [1979: 33–48]). It is less clear that 'If Germans are cruel then he is a Boche' fails to inherit the conventional implicature 'Germans are cruel' from 'He is a Boche.' These are matters for a more detailed account of pejoratives.

14. See also Hornsby (2001: 136). However, she registers significant doubts as to whether an appropriate content for (in present terms) the conventional implicature can be captured nonhomophonically. She also discusses the relation of pejoratives to accounts of evaluative terms by Hare and other moral philosophers. The bearing on such terms of the present essay is a large question.

well be true. When someone uses the word 'Boche,' one can legitimately ask him to withdraw the anti-German implication; but if it is clear, as it may well be, that he does believe that Germans are cruel, then it is hardly legitimate to ask him to withdraw the implication that he has that belief. Although one might try to persuade him to abandon the belief, and even succeed, the view at issue makes the implicature of his original remark be that he had the belief at the time of utterance, not that he has it now, which leaves no false implicature to withdraw. The false implicature is that Germans are cruel, not that the speaker believes that they are cruel. Perhaps the use of 'Boche' does also carry the additional implicature that the speaker believes that Germans are cruel, since a linguistically competent speaker who uses 'Boche' without believing that Germans are cruel is being insincere; but such a belief condition would be a by-product of the simple implicature that Germans are cruel, combined with the conversational norm of sincerity; it is not the source of what is most objectionable in the use of 'Boche.'

One might argue that the use of 'Boche' implies more than that Germans are cruel, namely, that the speaker knows that Germans are cruel, or even that it is common knowledge in the conversation that Germans are cruel. In asserting *p*, one implies in some sense that one knows *p*, although of course one does not thereby assert that one knows *p*, for one is in an epistemic position to assert *p* only if one knows *p*. Might implying *p* similarly generate the further implication that one is in an epistemic position to imply *p*? Such putative implications are not generated directly by the use of the pejorative word, but at best indirectly by the application of general principles of conversation to the direct implication; they will not be further discussed here.¹⁵

The conventional implicatures that words such as 'but' and 'Boche' generate are part of their meaning, in a broad sense of 'meaning.' If one is ignorant of them, one is at least partially ignorant of the meaning. An Englishman in Italy who thinks that 'ma' is synonymous with 'and' and 'e' with 'but' is mistaken, for 'e' is synonymous with 'and,' not with 'but,' and 'ma' is synonymous with 'but,' not with 'and.' Fully to understand a word,

15. On assertion, see Williamson (2000: 238–269). Evidently, an infinite regress looms if implying *p* involves implying that one knows *p* (in the same sense of 'imply'), for then one also implies that one knows that one knows *p*, and so ad infinitum. Arguably, all but the first few of these implications are false, even if 'know' is weakened to 'is in a position to know' (2000: 114–130). But no such regress flows from the principle that when one conventionally implies *p*, one conversationally implies that one knows *p*. In clause (2) of Grice's original definition of conversational implicature, 'the supposition that he is aware that, or thinks that, *q* is required in order to make his saying or making as if to say *p* (or doing so in *those* terms) consistent with this presumption [that he is observing the conversational maxims, or at least the Cooperative Principle]' (1989: 31), the qualification 'or doing so in *those* terms' may permit conventional implicatures and not just what is said to generate conversational implicatures. If not, an amendment along such lines would be consistent with the spirit of Grice's approach.

one must have some awareness, however inexplicit, of the conventional implicatures that it generates. In the case of 'Boche,' one might say, in Putnam's terminology, that cruelty is part of its associated *stereotype*; a stereotypical Boche is cruel. Putnam allows that stereotypes may be inaccurate; perhaps ferocity is part of the stereotype associated with the natural kind term 'gorilla,' although really gorillas are gentle. On his view, the stereotype for a word plays no direct role in determining its reference, but to be competent with the word one must have the stereotype (1975: 247–252). Since a competent speaker may know that the stereotype is inaccurate, to have the stereotype is not to believe that it is accurate; what one must be aware of is that it is the stereotype. Someone who understands 'Boche' may know that cruelty is an inaccurate part of the associated stereotype. The exact relation between conventional implicatures and stereotypes deserves further investigation, but we have a clear enough view for present purposes. What is most crucial is the separation of those aspects of meaning that contribute to truth-conditions from those that do not.¹⁶

'Boche' and 'German' have the same reference, so 'The Boche are the Germans' is true, but in using 'Boche' one implies that Germans are cruel. One can fully understand the word 'Boche' and know all that without being committed to the claim that Germans are cruel, for one can refuse to use the word 'Boche.' One is not obliged to utter every sentence that one knows to be true. One can know that a rule of inference is truth-preserving without using it. Nonxenophobic speakers of English may acknowledge that the rules of Boche-Introduction and Boche-Elimination* are truth-preserving yet still refuse to infer according to them. For one must *use* 'Boche' to infer according to those rules, whereas to classify them as truth-preserving is only to *mention* the word.¹⁷ Similarly, consider the Tarskian biconditional:

16. The argument in the text does not require conventional implicatures to form a homogeneous category, and therefore withstands recent arguments that they do not (Bach [1999], criticized by Carston [2002: 174–177]). Much recent discussion of conventional implicatures focuses on their role in organizing discourse, as in the case of 'but'; pejoratives typically play no such role. For a recent systematic and rigorous account of conventional implicature, see Potts (2005). Potts's application of his framework to expressives (153–193), for example, honorifics in Japanese, has points of similarity with this essay. However, to subsume the present account of 'Boche' under his framework, one must relax his constraint that no term has both truth-conditional ('at-issue') content and conventional implicature content. See also note 13 above.

17. Unlike Kaplan (2004), I do not want to replace truth-preservation as the standard of validity by 'information delimitation,' in the sense that any semantic information (including conventional implicatures) contained in the conclusion must already be contained in the premises. By Kaplan's standard, 'or'-introduction is an invalid rule, because ' $2 + 2 = 4$ or Lessing was a Boche' inherits the derogatory implication of 'Boche' and so contains semantic information not contained in ' $2 + 2 = 4$.' Similarly, the inference from 'This is not a camera' to 'This is not a Boche camera' counts as invalid by Kaplan's standard. Of course, Boche-Introduction counts as valid only in a less than purely logical sense of 'valid,' for 'Boche' and 'German' are not logical constants.

(T) 'Lessing was a Boche' is true in English if and only if Lessing was a Boche.

One can acknowledge that (T) is true in English but still refuse to utter it assertively, because one wants to avoid using 'Boche'.¹⁸

The present account has at least two further advantages over any account like Dummett's. First, it generalizes smoothly to adjectival uses of 'Boche.' In using 'Boche' as an adjective, one implies that Germans are cruel. Second, the account applies even to the narrow-minded community with the pejorative 'Boche' but no corresponding nonpejorative term. In using 'Boche,' they conventionally imply that Germans are cruel, although they can articulate the content of the implicature only by saying 'The Boche are cruel.'¹⁹

5. PEJORATIVES IN LANGUAGE AND THOUGHT

Pejoratives raise a further question about the nature of concepts. A vital feature of a pejorative *word* is that it carries a conventional implicature, but how could a pejorative *concept* do so? Conventional implicatures seem to arise from the communicative use of language. Although synonymous words in different languages carry the same conventional

18. See Barker (2003) for the problem that conventional implicature poses for asserting Tarskian biconditionals. However, the details of Barker's account do not seem well motivated. In acknowledging the truth of such biconditionals, the present truth-conditional approach is less reticent than that of Hornsby (2001).

19. Vagueness is another case in which willingness to infer by rules that are not truth-preserving has been treated as a precondition for having some defective concepts. Thus Dummett (1975) argues that observational predicates in natural language are governed by rules that infect the language with inconsistency: for example, to understand 'looks red,' one must be willing to apply a tolerance principle by which one can infer from 'x is visually indiscriminable from y' and 'x looks red' to 'y looks red,' which generates sorites paradoxes because visual indiscriminability is nontransitive. More recently, Roy Sorensen (2001) has argued that linguistic competence with vague terms involves willingness to make inferences such as that from '*n* seconds after noon is noonish' to '*n* + 1 seconds after noon is noonish,' which commits us to inconsistent conclusions by sorites reasoning (given our other commitments, such as 'Noon is noonish' and 'Midnight is not noonish'); he combines that view with an epistemic account of vagueness on which vague expressions have nontrivial classical extensions. By the present arguments, such claims about linguistic competence and concept possession are mistaken. An ordinary speaker of English who understands 'looks red' and 'noonish' and has the concepts *looks red* and *noonish* in the normal way but then rejects the relevant tolerance principles in the light of the sorites paradoxes does not thereby cease to understand those expressions or to have those concepts (perhaps she treats the premises of the tolerance principles as providing excellent defeasible evidence for their conclusions, an attitude that is less than Dummett and Sorensen require for competence). Even if the whole linguistic community abandons its supposed commitment to the tolerance principles, without stipulating any cutoff points, that would not make 'looks red' or 'noonish' any more precise, so speakers' acceptance of tolerance principles is in any case quite inessential to vagueness.

implicatures, how could they share those implicatures with a nonverbal concept? In the study of pejoratives, it is dangerous to give thought methodological priority over language by treating concepts as intrinsically nonlinguistic yet still capable of doubling as linguistic meanings. For that treatment is appropriate only if the false pejorative implicature is a feature of the concept expressed by 'Boche'; but then it is unclear how one could have the concept other than by understanding a word synonymous with 'Boche.' On such a view, the notion of having a concept seems to be parasitic on the notion of understanding a word, and language takes methodological priority over thought. If, on the other hand, the false implicature is not a feature of the concept expressed by 'Boche,' then presumably 'Boche' and 'German' express the same concept: since 'Boche' is not synonymous with 'German,' the meaning of a word is therefore not exhausted by the concept that it expresses, and the study of meanings cannot be subsumed under the study of concepts.²⁰ An adequate theory of concepts must resolve such questions.

It might be replied that since we can imagine someone who uses a pejorative concept in thought for which he has no corresponding word, the communicative use of language is inessential to pejorativeness. But that argument is too quick. We can imagine something like conversational implicature in thought, for example, when one thinks that there will be a nice view of the stars to imply to oneself that the roof of a dwelling has a large hole. Similarly, we can imagine something like conventional implicature in thought, for example, when one discriminates between one's own 'and' thoughts and 'but' thoughts. However, such examples show only that communicative phenomena can occur even in silent communication with oneself. They are cases in which one manipulates the rhetorical effects of one's own thoughts on oneself. A significant motive for the methodological switch from the level of language to the level of thought was precisely to avoid such complexities of the process of communication in the most elementary examples. When those complexities are reintroduced at the level of thought, the putative gain in simplicity disappears. If a straightforward inferentialist account of pejoratives were correct, then pejorative concepts could occur at a quite primitive level of thought, since the corresponding inferential roles could be realized at that level. By contrast, if the present account is correct, then pejorativeness is a more sophisticated phenomenon than the inferentialist account suggests.²¹

20. If concepts are thought constituents, and *Boche* and *German* are the same concept, then in judging *Germans are German* one simultaneously judges *Germans are Boche*, however much one's reactions discriminate between the sentences 'Germans are German' and 'Germans are Boche.' Compare: if *furze* and *gorse* are the same concept, then in judging *furze is furze* one simultaneously judges *furze is gorse*, however much one's reactions discriminate between the sentences 'Furze is furze' and 'Furze is gorse.'

21. This essay elaborates ideas in one part of Williamson 2003, which together with Boghossian 2003 constituted the symposium 'Blind Reasoning' at the 2003 Joint Session of

ACKNOWLEDGMENTS. This essay discusses some of the same phenomena as the fascinating Kaplan 2004, and reaches conclusions that are in some respects quite similar (a difference concerning the notion of validity is noted in n. 17). One sign of my general debt to David Kaplan is the difficulty that I have in writing a paper without citing Kaplan 1989, a difficulty that I share with many other philosophers of language. I first encountered his work as an undergraduate, when I read Kaplan 1969, and was immediately impressed by his intellectual fertility, his rigor, and his playfulness. Opponents of the scientific spirit in philosophy often associate it with humourless severity, sterility, and indifference to nuance and aesthetic value. David is a wonderful counterexample. Playfulness is one of the best antidotes to that toxin for the scientific spirit, the desire for salvation from philosophy. Precision forces one to respect the subtle distinctions that free-flowing 'humanistic' prose pours indifferently over. Rigor provides the constraints that distinguish creativity from arbitrary variation. By precedent rather than precept, logic teaches the value of elegance and a sense of form, even in the search for truth.

REFERENCES

- Bach, K. 1999. "The Myth of Conventional Implicature." *Linguistics and Philosophy* 22, 327–366.
- Barker, S. 2003. "Truth and Conventional Implicature." *Mind* 112, 1–33.
- Belnap, N. 1962. "Tonk, Plonk and Plink." *Analysis* 22, 130–134.
- Boghossian, P. 1989. "The Rule-Following Considerations." *Mind* 98, 507–549.
- Boghossian, P. 2002. "How Are Objective Epistemic Reasons Possible?" In J. Bermudez and A. Millar, eds., *Reason and Nature*. Oxford: Oxford University Press.
- Boghossian, P. 2003. "Blind Reasoning." *Aristotelian Society*, suppl. vol. 77, 225–248.
- Brandom, R. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, Mass.: Harvard University Press.
- Brandom, R. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass.: Harvard University Press.
- Carston, R. 2002. *Thoughts and Utterances: The Pragmatics of Explicit Communication*. Oxford: Blackwell.
- Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.

the Aristotelian Society and Mind Association in Belfast. I thank Paul Boghossian and the audience there for valuable discussion. Subsequent to the writing of that paper, the present material was further developed in talks at the universities of Saint Andrews, Bucharest, Pittsburgh, and UCLA, and for conferences at the Philosophical Institute in Zagreb, Yale University, and the University of Stockholm. I thank the audiences there, not least David Kaplan himself at Yale and UCLA, for helpful reactions. Malte Dahlgrün and John Macfarlane provided useful written comments, and I have benefited from Owen Greenhall's expertise in this area. Older debts are to Jason Stanley and participants in classes at Oxford.

- Dummett, M. 1975. "Wang's Paradox." *Synthese* 30, 301–324.
- Dummett, M. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.
- Evans, G. 1982. *The Varieties of Reference*. Ed. J. McDowell. Oxford: Clarendon Press.
- Frege, G. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Nebert. Trans. S. Bauer-Mengelberg in J. van Heijenoort, ed., *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, Mass.: Harvard University Press, 1967.
- Frege, G. 1979. *Posthumous Writings*. Ed. H. Hermes, F. Kambartel, and F. Kaulbach. Trans. P. Long and R. White. Oxford: Basil Blackwell.
- Grice, H. P. 1961. "The Causal Theory of Perception." *Proceedings of the Aristotelian Society*, suppl. vol. 35, 121–152.
- Grice, H. P. 1989. *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.
- Horn, L. 1989. *A Natural History of Negation*. Chicago: University of Chicago Press.
- Hornsby, J. 2001. "Meaning and Uselessness: How to Think about Derogatory Words." In P. A. French and H. K. Wettstein, eds., *Midwest Studies in Philosophy XXV: Figurative Language*. Oxford: Blackwell.
- Kaplan, D. 1969. "Quantifying In." In D. Davidson and J. Hintikka, eds., *Words and Objections: Essays on the Philosophy of W. V. Quine*. Dordrecht: Reidel.
- Kaplan, D. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals." In J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. New York: Oxford University Press.
- Kaplan, D. 2004. "The Meaning of Ouch and Oops: Explorations in the Theory of *Meaning as Use*." Typescript.
- Karttunen, L., and S. Peters. 1979. "Conventional Implicature." In C.-K. Oh and D. A. Dinneen, eds., *Syntax and Semantics, 11*. New York: Academic Press.
- Kripke, S. 1979. "A Puzzle about Belief." In A. Margalit, ed., *Meaning and Use*. Dordrecht: Reidel.
- McGee, V. 2000. "'Everything'" In G. Sher and R. Tieszen, eds., *Between Logic and Intuition*. Cambridge: Cambridge University Press.
- Peacocke, C. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.
- Peacocke, C. 1998. "Implicit Conceptions, Understanding and Rationality." In E. Villanueva ed., *Philosophical Issues, 8: Concepts*. Atascadero, Calif.: Ridgeview.
- Potts, C. 2005. *The Logic of Conversational Implicatures*. Oxford: Oxford University Press.
- Prior, A. N. 1960. "The Runabout Inference-Ticket." *Analysis* 21, 38–39.
- Putnam, H. 1975. "The Meaning of 'Meaning'." In H. Putnam, *Mind, Language and Reality*. Cambridge: Cambridge University Press.
- Rumfitt, I. 2000. "'Yes' and 'No.'" *Mind* 109, 781–823.
- Sorensen, R. 2001. *Vagueness and Contradiction*. Oxford: Clarendon Press.
- Stalnaker, R. 1999. *Context and Content*. Oxford: Oxford University Press.
- Stanley, J., and T. Williamson. 2001. "Knowing How." *Journal of Philosophy* 98, 411–444.
- Strawson, P. F. 1952. *Introduction to Logical Theory*. London: Methuen.

- Williamson, T. 1987/88. "Equivocation and Existence." *Proceedings of the Aristotelian Society* 87, 109–127.
- Williamson, T. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. 2003. "Understanding and Inference." *Aristotelian Society*, suppl. vol. 77, 249–293.

CHAPTER 10

Frege's Hierarchy

A Puzzle

Christopher Peacocke

To my great good fortune, David Kaplan visited Harvard for the fall of 1971, during a memorable year I spent there as a Kennedy Scholar. I had studied his writings, but nothing could quite prepare you for the man himself. There is his energy, the love of the issues, the wit, the thorough formal mastery, the sharp and inquiring eye for the linguistic and mental phenomena, the tremendous generosity with his time in individual philosophical discussions. Beyond these personal characteristics, two things stuck with me about how one should aim to do the subject. When you have a philosophical conversation with David, for much of the time his stance in the exchange is actually one that adopts *your* position. He elaborates it, expounds its connections to other issues, makes you see things in your view you never suspected were there—and thereby shows you are completely wrong. It is an exercise of philosophical empathy, a way of working through to the truth that is available only to someone who has engaged in deep reflection on what motivates the views with which he disagrees.

The other quality that so impressed me is the realization in his work of the possibilities of bringing the discipline, powers, and explicitness of the apparatus of logic and semantics to philosophical problems and issues of which it is not at all obvious in advance that they are amenable to such techniques. This quality is already present in one of his earliest works (with Richard Montague), the elegant paper “A Paradox Regained.”¹ It is prominent in a different form in “Quantifying In,” a paper unsurpassed in its resourcefulness and fertility in discriminating and characterizing mental phenomena, and their various relations to the world, by means of the classical Fregean semantical apparatus.² “Quantifying In” has by now

1. D. Kaplan and R. Montague, “A Paradox Regained,” *Notre Dame Journal of Formal Logic* 1 (1960): 79–90.

2. D. Kaplan, “Quantifying In,” in *Words and Objections: Essays on the Work of W. V. Quine*, ed. D. Davidson and J. Hintikka (Dordrecht: Reidel, 1969).

been studied by the philosophical community as closely as the works it discusses. The richness of this paper is not yet exhausted, and I will be suggesting a new application of one of its ideas in what follows.

I will be discussing the Fregean hierarchy of senses. This is a topic on which many of the leading writers on Fregean sense of the past sixty years have published, including, in order of publication, Church, Kaplan himself, Davidson, Dummett, Burge, and Terence Parsons.³ The interest of the issues is far from purely technical. A resolution of issues about the hierarchy of senses turns on the resolution of a series of issues of much wider significance in the theory of thought, in epistemology, in metaphysics, and in the philosophy of mind. The acceptability and nature of a Fregean hierarchy of senses involves the correct way of conceiving of the relation between sense and reference. It involves some conception of how thinkers are capable of thinking of abstract objects like senses and Thoughts. (I use 'Thoughts' with an initial capital for the propositional intentional contents, Fregean *Gedanken*, to distinguish them from mental events, including thinkings.) Since the hierarchy in turn bears upon issues about thinking about Thoughts, resolutions of these issues also have ramifications in the philosophy of mind for what is involved in the ascription of attitudes to oneself and to others. Correspondingly, evidence drawn from these various areas is pertinent to the assessment of treatments of the Fregean hierarchy. I will try to draw out some of these connections in what follows.⁴

I begin with the puzzle.

1. THE PUZZLE

When Karl believes

(1) Bush is powerful

3. A. Church, "A Formulation of the Logic of Sense and Denotation" in *Structure, Method and Meaning: Essays in Honor of Henry M. Sheffer*, ed. P. Henle, H. Kallen, and S. Langer (New York: Liberal Arts Press, 1951); D. Kaplan, *Foundations of Intensional Logic* (UCLA Dissertation, 1964: Ann Arbor, University Microfilms); D. Davidson, "Theories of Meaning and Learnable Languages" and "On Saying That," papers from 1964 and 1969, respectively, both reprinted in his *Inquiries into Truth and Interpretation*, 2nd ed. (Oxford: Oxford University Press, 2001); M. Dummett, *Frege: Philosophy of Language* (London: Duckworth, 1973); T. Burge, "Frege and the Hierarchy," *Synthese* 40 (1979): 265–281, reprinted with extensive further discussion in a Postscript in his *Truth, Thought, Reason: Essays on Frege* (Oxford: Oxford University Press, 2005); T. Parsons, "Frege's Hierarchies of Indirect Sense and the Paradox of Analysis," *Midwest Studies in Philosophy* 6 (1981): 37–57.

4. A version of this essay was submitted for this volume in 2005. Paolo Leonardi very kindly allowed me to take advantage of delays in completing his collection and to submit a revised version in 2007. Special thanks to Tyler Burge for much intensive and illuminating discussion and correspondence over the past twelve years on the issues of this essay, and to Richard Heck for valuable comments and queries on the earlier draft. There is much more to be said on the relation of the present essay both to Burge's and to my other published contributions on these issues, but it has to be left for another occasion if this essay is to remain relatively self-contained.

he believes something about the world: about a man and his power. In so believing, Karl thinks about Bush in a particular way, a way that we can write $\langle \text{Bush} \rangle$. Equally we can write $\langle \text{is powerful} \rangle$ for Karl's way of thinking of the property of being powerful. If, unlike Frege but in accordance with contemporary usage, we use 'concept' for anything that can be a constituent of a Thought, we can quite generally write $\langle A \rangle$ for the concept that is the sense of the expression A (if it has a sense). Under the classical Fregean treatment, the sentence

(2) Karl believes that Bush is powerful

states that Karl stands in a certain relation to a Thought built up from concepts. If we use " $\wedge$ " as a symbol for predicational combination of concepts, then according to the classical Fregean account, (2) states that

(3) Bel (Karl, $\langle \text{Bush} \rangle \wedge \langle \text{is powerful} \rangle$).

If John believes what (2) states, that is, if John believes that Karl believes that Bush is powerful, then according to the classical, hierarchical Fregean account, the following is true:

(4) Bel (John, $\langle \text{Bel} \rangle \wedge \langle \text{Karl} \rangle \wedge \langle \langle \text{Bush} \rangle \wedge \langle \text{is powerful} \rangle \rangle$).

(I simplify the symbolic formulation for legibility.)

I start with a tiny question: What is the relation between Karl's way of thinking about Bush, and John's way of thinking about Karl's way of thinking about Bush? Are they identical? More generally, is $\langle A \rangle$ identical with $\langle \langle A \rangle \rangle$? If they are not identical, what is the relation between them?

Despite its brevity, the question of whether $\langle A \rangle = \langle \langle A \rangle \rangle$ has extensive ramifications and theoretical significance. Some have held that Fregean theories are unable to give any satisfactory answer to this question and have drawn the conclusion that the notion of sense is not suitable for giving an account of the content of propositional attitudes.⁵ But addressing the question goes beyond the issue of the theoretical utility of the notion of sense. Any answer to the question must draw on a conception of the relations between two intertwined abilities we possess: the ability to think about the world in certain ways, and the ability to think about how we and others are thinking about the world.

It is intuitively plausible that the ability to think about any Thought as the Thought that p requires the ability to employ that same Thought p in first-order thinking about the world. If this is so, why is it so? We need a general theoretical explanation of this intuitive truth. How, in the enterprise of philosophical explanation, are we to conceive of the connection between the complex of relations required to think about something in

5. Davidson, "Theories of Meaning and Learnable Languages."

the world in a given way, on the one hand, and the ability to think about that very way itself, on the other?

The question of whether $\langle A \rangle = \langle \langle A \rangle \rangle$ has, more specifically, been a puzzle because there seem to be compelling reasons for a negative answer, and there seem to be compelling reasons for a positive answer. I group the considerations in support of the distinctness of $\langle A \rangle$ and $\langle \langle A \rangle \rangle$ under the heading "Thesis." I group the considerations in support of the identity of $\langle A \rangle$ and $\langle \langle A \rangle \rangle$ under the heading "Antithesis."

Thesis

$\langle A \rangle$ cannot be $\langle \langle A \rangle \rangle$ because:

(a) Each concept is individuated by the fundamental condition for something to be its reference. This is a Fregean conception, and its importance has long been emphasized by Dummett.⁶ The conception applies just as much to concepts of concepts as it does to concepts of ordinary objects. The condition for being the reference of a first-level concept $\langle A \rangle$ must, together with the world, determine something at the level of reference (an object, in the case of singular concepts). By contrast, the condition for being the reference of a concept $\langle \langle A \rangle \rangle$ must, together with the world, determine a concept, rather than something at the level of first-level references. Since these conditions determine different references, given the way the world is, they must be distinct. Hence, given the individuation of a concept by the fundamental condition for being its reference, these conditions correspond to distinct concepts. So $\langle \langle A \rangle \rangle$ is not $\langle A \rangle$. We can call this 'the condition-for-reference argument'.

(b) A concept is, as Evans said, a way of thinking of something.⁷ A way of thinking cannot be both a way of thinking of a thing and also a way of thinking of a concept of that thing. Hence $\langle \langle A \rangle \rangle$ is not $\langle A \rangle$. This is 'the way-of-thinking argument'. This intuitive argument was presented to me in correspondence by Tyler Burge and features in his "Postscript to 'Frege and the Hierarchy.'"⁸

(c) Concepts, according to some (but by no means all) theorists of concepts, are individuated by their possession conditions.⁹ Whatever the condition a thinker must meet to possess the concept $\langle A \rangle$, it involves

6. See Dummett, *Frege: Philosophy of Language*, 93ff.; and *The Interpretation of Frege's Philosophy* (London: Duckworth, 1981), esp. 42–45. This Dummettian point about sense-reference relations is being used in this first consideration for the Thesis in support of a conclusion about the hierarchy of senses that Dummett elsewhere rejects. For more general arguments in support of a conception of a concept as individuated by the fundamental condition for something to be its reference, see my *Truly Understood* (Oxford: Oxford University Press, 2008), chaps. 1–4.

7. G. Evans, *The Varieties of Reference* (Oxford: Oxford University Press, 1982), chap. 1.

8. T. Burge, *Truth, Thought, Reason*.

9. C. Peacocke, *A Study of Concepts* (Cambridge, Mass.: MIT Press, 1992), chaps. 1 and 2, and the less stringent conception, which still respects the idea that concepts are individuated

only his ability to think about the world. But to possess $\langle\langle A \rangle\rangle$ the thinker must have the ability to think about concepts. So $\langle A \rangle$ and $\langle\langle A \rangle\rangle$ have distinct possession conditions and hence are distinct concepts. This is 'the possession-condition argument'.

Under some approaches to concepts, namely, those that accept all the characterizations of concepts given in these three arguments, these three arguments are different perspectives on a single underlying idea. The idea underlying these arguments for the Thesis is that the very essence of concepts, whose three characterizations in terms of reference-conditions, ways of thinking, or possession-conditions can be shown to be equivalent, is incompatible with a given concept having such different things as an object and a distinct concept as its reference.¹⁰

This concludes the arguments for the Thesis.

Antithesis

$\langle A \rangle$ must be $\langle\langle A \rangle\rangle$ because:

(a) When you understand 'believes' (for instance) and any arbitrary term a and sentence s , you are thereby in a position to understand the sentence ' a believes that s '. The simplest explanation for this, the argument runs, is that $\langle\langle A \rangle\rangle$ is the very same concept as $\langle A \rangle$. This is the argument from understanding propositional attitude contexts.¹¹

(b) If $\langle A \rangle$ were distinct from $\langle\langle A \rangle\rangle$, there would be an infinite hierarchy of concepts. Concepts and concepts of concepts are in a one-many relation: For anything at all, including concepts, there are many concepts of it. There would then be no explanation of how we can grasp this hierarchy, for, it is said, we have been given no principle that generates $\langle\langle A \rangle\rangle$ from $\langle A \rangle$. This was Davidson's argument in "Theories of Meaning and Learnable Languages" and "On Saying That." Yet if the Fregean hierarchy exists, there would have to be some principle generating the members of a higher level from the members of the next level down, since we can understand arbitrary long embeddings of propositional attitudes. Since there is, the argument runs, no such generating principle, if we are going to use the notion of a concept at all, we had better hold that $\langle A \rangle = \langle\langle A \rangle\rangle$. This is the argument against a hierarchy from the absence of a generating principle.

(c) There are many inferences whose validity is, apparently, most easily explained by the hypothesis that $\langle A \rangle = \langle\langle A \rangle\rangle$. From

by their possession-conditions, that grasp of a concept is tacit knowledge of its fundamental rule of reference.

10. For a defense of the equivalence of these three conceptions of sense, see my "Concepts without Words," in *Language, Thought and Logic: Essays in Honour of Michael Dummett*, ed. R. Heck (Oxford: Oxford University Press, 1997).

11. See my "Entitlement, Self-Knowledge and Conceptual Redeployment," *Proceedings of the Aristotelian Society* 96 (1996): 117–158.

(5) John believes that Karl believes that Bush is powerful

it seems to follow that

(6) There is some proposition $\langle p \rangle$ such that John believes that Karl believes that $\langle p \rangle$, and $\langle p \rangle$ is the proposition that Bush is powerful.

Similarly, from (5) and (7)

(7) It is true that Bush is powerful

it certainly seems to follow that

(8) Something John believes Karl believes is true.

If, quite generally, $\langle A \rangle$ is $\langle \langle A \rangle \rangle$, and in particular $\langle \text{Bush is powerful} \rangle = \langle \langle \text{Bush is powerful} \rangle \rangle$, both these arguments are validated, without any supplementary premises or principles.¹² According to this position, the clause 'that Bush is powerful' always refers to the same thing, however deeply it is embedded. But if the Thought $\langle \text{Bush is powerful} \rangle$ were distinct from the Thought $\langle \langle \text{Bush is powerful} \rangle \rangle$, then the arguments would need additional premises for validity, which apparently they do not. In natural language, we apparently treat a given clause of the form 'that p ' as having the same reference whether it is embedded in one or arbitrarily many operators, be they operators of propositional attitudes or such operators as 'It is true that', 'It is probable that', and the like. The phenomenon is not peculiar to English. It occurs in other natural languages, including French and German. The challenge is rather to find a natural language that does behave in accordance with the Fregean hierarchy of concepts. I label this 'the argument from inferences'.

That concludes the arguments for the Antithesis.

2. A PROPOSAL

If we could show how the canonical concept of a concept s is uniquely determined by s itself, and could explain the principle by which it is determined, we would meet the need for a generating principle.

Now there are special cases in which a concept of something is determined by the thing itself. One such class of special cases is that of certain concepts of certain abstract objects. For instance, a person may think of an object under the very condition that individuates that object. The number 0 itself is individuated by its being the unique number n of which it holds that necessarily there are n things that are F iff $\sim \exists x(Fx)$. That fact about the number 0 is a truth of metaphysics, about the nature of numbers. But there is a corresponding truth at the level of thought. To

12. This was an argument I developed in "Entitlement, Self-Knowledge and Conceptual Redeployment."

think of a number as 0 is to have tacit knowledge that it is the unique number n such that necessarily there are n things that are F iff $\sim\exists x(Fx)$. An analogous point can be made about thinking of a number as 1. The number 1 is individuated by the fact that it is the unique number n such that necessarily there are n things that are F iff $\exists x\forall y(Fx \leftrightarrow y = x)$. Thinking of a number as 1 involves tacit knowledge of that same individuating condition. Here we have examples at the level of individuals of a phenomenon that also occurs for properties and relations: that of thinking of property or relation in a way fixed by what makes the property or relation the property or relation it is.¹³

So, there are some cases in which a canonical concept of an object is determined by the object itself. Does anything analogous hold for canonical concepts of concepts, for concepts of concepts and Thoughts?

I argue that it does. I use the notation ' $\text{can}(F)$ ' for the canonical concept of the concept F . At the level of metaphysics, $\text{can}(F)$, like any other concept, is individuated by its fundamental reference rule, in accordance with a broadly Fregean conception of concepts. In the case in which F is a first-level concept, the fundamental reference rule for $\text{can}(F)$ will state a condition for an arbitrary concept to fall under $\text{can}(F)$. A plausible first attempt to state that condition is this:

- (*) For an arbitrary concept C to fall under $\text{can}(F)$ is for the fundamental condition for something to fall under C to be that it meets condition R , where R is in fact the fundamental reference rule for the concept F .

(*) makes clear that the canonical concept of F , $\text{can}(F)$, unlike other ways of thinking of a concept, is determined by the concept itself. This is so because the fundamental condition for something to fall under the concept F is determined by the concept itself, rather than some mode of presentation of the concept F .

In parallel with the numerical example, we can then say that to think of a concept under the canonical concept of F , under $\text{can}(F)$, is to have tacit knowledge of what (*) states. I call this claim the 'Leverage Account' of what it is to be employing $\text{can}(F)$ in thought. It leverages the materials of a condition for something to be a concept F of one level into the materials of an account of what it is to be a particular concept, the canonical concept of F , at a higher level in the Fregean hierarchy of concepts.

13. For more discussion of some other examples, see my paper "Explaining the *A Priori*," in *New Essays on the A Priori*, ed. P. Boghossian and C. Peacocke (Oxford: Oxford University Press, 2000). Issues concerning thought about numbers are very delicate, and there are close limits to the extent to which the points in the text can be generalized to other, even canonical, ways of thinking of natural numbers. To think of something as 5, for instance, does not necessarily involve thinking of it as the successor of 4. When you see, without counting, that there are five mugs on the table, it is a further step, and a further thought, that the number of mugs there is the successor of 4. Also, thinking of a number as 1001 in binary notation is different from thinking of it as 9 in Arabic notation.

Like explicit knowledge, tacit knowledge may be merely partial. Intuitively, a thinker can think of a concept as the concept *turbulent* even if she has only a partial grasp of the first-level concept *turbulent*. Such a thinker has only a partial tacit knowledge of the reference rule for the concept *turbulent* and so has only a partial tacit knowledge of the reference rule for the canonical concept of the concept *turbulent*.

Under the Leverage Account so conceived, it is a distinctive feature of the canonical concept of F that a thinker cannot possess that canonical concept unless he has at least a partial grasp of the concept F itself. This is part of the answer to the question of how the ability to think of the world is intertwined with the ability to think about one's own and others' thought about the world. In possessing this distinctive feature, the canonical concept differs from indefinitely many other concepts of the concept F. Such concepts of concepts as *Einstein's favorite concept*, or *the legal concept specified on the title page of such-and-such book*, are concepts that a thinker can possess without possessing the concepts to which they refer, and without knowing which concepts they are. By contrast, someone who possesses the concept *the concept number* must have at least a partial grasp of the concept *number* and know which concept *the concept number* refers to. In certain respects, then, as Jerry Fodor once remarked to me, applying the functional concept *the concept of...* to a concept is analogous to applying quotation marks to an expression. If you understand quotation marks, and perceive an expression of the form '___', you know what that whole expression refers to.

Another consequence of this conception is that whatever contributes to the individuation of the concept F thereby also contributes to the individuation of the concept *can*(F). In particular, if F is externally individuated, the condition R will involve external conditions, and correspondingly *can*(F) will also be externally individuated. This fact has an epistemological significance in the philosophy of mind I will take up in section 5.

Attribution of tacit knowledge of some content is justified by what attribution of that tacit knowledge explains. What does tacit knowledge of the fundamental reference rule (*) explain?

One very straightforward class of transitions in thought explained by tacit knowledge of (*) is the thinker's willingness to move from

a is F

to

a falls under the concept of F;

and conversely. Tacit knowledge of (*) explains the willingness to make this transition, since if something is F, it evidently falls under a concept the fundamental condition to fall under which is the same as the fundamental condition for an object to fall under F (and conversely). The

fundamental reference rule (*) in effect correspondingly lifts reasons for judging that something is *F* up one level of the Fregean hierarchy to reasons for judging that something falls under *can(F)*.

A philosophically more interesting engagement of tacit knowledge of (*) arises in its interaction with mental states in whose content the concept *F* itself features (rather than any mode of presentation of the concept *F*). Consider a thinker who makes a transition from his action-awareness of

judging *Fa*

to the self-ascription

I judge that *Fa*.

In attributing this content to his judgment, he is thinking of the concepts *a* and *F* under their respective canonical modes of presentation. If the thinker has a sensitivity to which of his mental states contain the concept *F* rather than some other concept, this sensitivity can be combined with tacit knowledge of (*) to explain the thinker's ability to make such self-ascriptions—and, in the right circumstances, thereby to gain knowledge of the content of his mental states. The characterization of the states to which the thinker has to be sensitive does not require him already to think of the concepts *a* and *F* under their canonical concepts. The sensitivity is just to the presence of *a* and *F* themselves in the content of the mental state. Actually, the same applies to the concept of judging too. Grasp of this concept involves an ability to be sensitive to the distinction between those events that are judgments and those that are not. The capacity of a thinker to be sensitive to the presence or absence of some feature itself, and the presence of this capacity in the account of mastery of some concept of that feature itself, seems to me to be a widespread phenomenon in the theory of concepts. For anyone who believes in the existence of a level of nonconceptual representational content in perception, and holds that observational concepts are individuated by their relations to such contents, this phenomenon of sensitivity to a feature without prior conceptualization thereof is present even in that basic case.¹⁴

It is very plausible that part of the subpersonal explanation of how it is that a thinker is able to enjoy such content-sensitivity in his grasp of the canonical concept of a concept is that the subpersonal realizations of the relevant mental states to which he is sensitive contain some representations in a subpersonal, Fodorian language of thought, structured representations that have the Thought in question as its assigned concept. I agree with those who say that it is hard to see how there can be any other explanation of all the phenomena in which mental states with

14. Such a treatment is given in chapter 3 of *A Study of Concepts*.

content are implicated.¹⁵ It matters, however, that the sensitivity mentioned in the account of the ability to employ a canonical concept of a Thought does not merely involve subpersonal sensitivity to the occurrence of a particular formula in the subpersonal language of thought. By itself, such a subpersonal sensitivity is neither necessary nor sufficient for the capacity with which we are concerned. It is not sufficient, because that sensitivity in a subject whose subpersonal symbols meant something different would not be the sensitivity we require, which is a sensitivity to a concept itself in the content of a mental state. It is not necessary, because any other symbol for the same Thought would equally serve. The particular subpersonal structured symbol whose sense is a given Thought is implicated in the realization of our subject's capacities, but it is not constitutive of them. The capacities are characterized in content-involving terms. The sensitivity to the occurrence of mental states containing a given content will be a sensitivity to a state whose content is individuated in part by its relations to other states, and, very often, in part by its relations to the subject's environment. The occurrence of particular expressions in the language of thought cannot by itself ensure the obtaining of these relations. The states of affairs and events in whose empirical explanation canonical concepts of concepts are involved will also be individuated by their relations to other states, and (often) to the environment. Symbolic forms are by themselves inadequate to explain these effects too. These should be seen as familiar points about explanation by content-involving states, personal or subpersonal.¹⁶ They are simply applied here in the context of thought about concepts and Thoughts.

The fact that someone possesses a given concept F makes it possible for him to think of it in a distinctive way, as the concept F, a way available to him only because he possesses the concept F. In this respect, canonical ways of thinking of particular concepts resemble ways of thinking of mental states and events that are available to a thinker only because he knows what it is like to enjoy such states and events. Since we have a recognitional capacity for events that are our own pains and our own seeings, for example, we have ways of thinking of them for which it is an open question what their role is in a thinker's psychological economy. To think of an event as a pain, or as a seeing, is not thereby to think of it, even tacitly, as whatever plays such-and-such role in a thinker's psychological economy. In effect, I have just been arguing a similar case for the rather different category of canonical thought about a given concept F, canonical thought made available by the thinker's own grasp of the concept F itself. Though concepts are quite different sorts of things

15. J. Fodor, *The Language of Thought* (New York: Crowell, 1975); M. Davies, "Concepts, Connectionism and the Language of Thought," in *Philosophy and Connectionist Theory*, ed. D. Rumelhart, W. Ramsey, and S. Stich (Hillsdale, N.J.: Erlbaum, 1991).

16. J. Hornsby, "Physicalist Thinking and Conceptions of Behaviour," in *Subject, Thought and Context*, ed. P. Pettit and J. McDowell (Oxford: Oxford University Press, 1986); and my "Externalist Explanation," *Proceedings of the Aristotelian Society* 93 (1993): 203–230.

from mental events and conscious states, it is true both of concepts and of conscious states and events that we stand in a special relation to them that makes available canonical ways of thinking of them that cannot be identified with an instance of the form 'whatever has such-and-such role'. The same point applies *pari passu* for canonical ways of thinking of meanings, in the cases in which meaning can be identified with concepts as understood here.

To summarize: On the Leverage Account, for each concept F , there is a canonical way of thinking $\text{can}(F)$ of the concept F . This canonical way is uniquely fixed, under that account, by what individuates the concept F itself.

Under the Leverage Account, the subject's ability to think about concepts and Thoughts at all is explanatorily posterior, as a constitutive matter, to his ability to employ first-level (or lower-level) concepts and Thoughts.

The Leverage Account can be iterated up the hierarchy of concepts. We can say that

(**) For an arbitrary concept C to fall under the concept $\text{can}(\text{can}(F))$ is for C to be such that: the fundamental condition for something to fall under C is R , which is in fact the fundamental condition for an object (entity) to fall under $\text{can}(F)$.

Just as is the case one level down, the canonical concept $\text{can}(\text{can}(F))$ is individuated by the concept $\text{can}(F)$ itself on this account, and not by some mode of presentation of $\text{can}(F)$. It is clear that this approach can be iterated up the hierarchy to any finite level. It gives a generating principle for fixing the canonical concept of a concept from the latter concept itself. Tacit grasp of this generating principle will allow thinkers to understand arbitrary embeddings of propositional-attitude verbs.

3. HOW THE SOLUTION TREATS THE CONSTRAINTS THAT GENERATE THE PUZZLE

The Leverage Account accepts all of the arguments for the Thesis that $\langle\langle A \rangle\rangle$ must be distinct from $\langle A \rangle$.

The condition-for-reference argument could be expanded, in the presence of the Leverage Account. Consider the case in which F is a first-level concept, true or false of objects. Suppose, following Frege, that we take the reference of a concept to be a function whose range consists of truth-values. Then under the Leverage Account, the condition for something to be mapped by the reference of $\text{can}(F)$ to the True certainly requires it to be a concept. This is certainly distinct from the condition required for something to be mapped to the True by the reference of F , which will (on the Fregean view) require it to be an object, meeting whatever condition is determined by the first-level concept F . The conditions required for

something to be the reference of $\text{can}(F)$ are distinct from the conditions for something to be the reference of the first-level concept F , just as the condition-for-reference argument maintained. The same argument could be reproduced further up the hierarchy.

The Leverage Account can also endorse the second, intuitive, argument for the Thesis in terms of ways of thinking. If a way of thinking is individuated by the fundamental condition for something to be its reference, a way of thinking of a concept cannot be the same as a way of thinking of something that is not a concept. The ways of thinking must be individuated by distinct conditions for things to be their respective references.

The third argument, from the distinctness of the possession conditions for $\langle\langle A \rangle\rangle$ and $\langle A \rangle$, is also underwritten and validated on the Leverage Account. The possession condition for the canonical concept of a concept involves tacit knowledge, of $(*)$, that involves the ability to think about concepts. The possession condition for the concept so thought about does not involve relations to contents but only relations to the world itself (or possibly to lower-level attitudes and Thoughts). So the possession conditions for $\langle\langle A \rangle\rangle$ and $\langle A \rangle$ are distinct, as the third argument for the Thesis maintained.

The Leverage Account as formulated already has the resources for responding head-on to the first of the two arguments for the Antithesis. The first of those arguments, the appeal to understanding propositional attitude contexts, cited the fact that when you understand “believes” (for instance) and any arbitrary term a and sentence s , you are thereby in a position to understand the sentence ‘ a believes that s ’. Contrary to the argument in the Antithesis, however, under the identity-constitutive view, the explanation of this fact is not that $\langle\langle A \rangle\rangle = \langle A \rangle$. It is simply that any concept uniquely determines the canonical concept of that concept, in the way described in the Leverage Account. Because thinkers are able to employ a canonical concept of a concept that is uniquely determined by that concept, as long as they have the general ability to move from a concept to its canonical concept, they will be able to understand any sentence of the form “ a believes that s ” as soon as they can understand its constituents. Understanding the propositional attitude form just involves appreciation that if s expresses the Thought $\langle p \rangle$, then “ a believes that s ” expresses the Thought $\langle \text{Bel} \rangle^{\wedge} \langle a \rangle^{\wedge} \langle \langle p \rangle \rangle$.

The Leverage Account supplies the generating principle for the hierarchy demanded in the second argument for the Antithesis. So for the first two arguments marshaled in the case for the Antithesis, the Leverage Account can explain why the considerations they adduce, though correct in themselves, do not support the thesis that $\langle\langle A \rangle\rangle$ is identical with $\langle A \rangle$.

The status of the third argument for the Antithesis, from the validity of inferences, is more complex. The two transitions from (5) to (6), and from (5) and (7) to (8), are validated under the Leverage Account if we add the premise, uncontroversial on the account, that

$\langle\langle p \rangle\rangle$ refers to the Thought $\langle p \rangle$.

So, when we say (6) above follows from (5), (6) is understood as meaning that

(9) There is some Thought $\langle p \rangle$ such that there is a concept $\langle\langle p \rangle\rangle$ that refers to $\langle p \rangle$, and such that John believes that Karl believes $\langle\langle p \rangle\rangle$, and $\langle p \rangle$ is the Thought that Bush is powerful.

I have abbreviated some formalities for simplicity.

This, however, is far from an adequate answer to the third argument for the Antithesis. The point of that argument was that *no* additional premise, however obvious, should be needed—the language behaves as if ‘that p ’ refers to the same thing, however embedded, functioning as a complex term for the same Thought wherever it occurs. This was precisely the consideration that was influencing some of my earlier writings on this topic.¹⁷ So this challenge has not yet been fully addressed.

A first step toward addressing the challenge involves a consideration about canonical concepts of concepts more generally. It is very plausible that if s is a concept of an abstract object x that is determined by what individuates x , as in the example of the numbers 0 and 1 we considered above, then what Quine in “Quantifiers and Propositional Attitudes” called ‘exportation’ is valid for it.¹⁸ That is, if someone believes $F(s)$, then she believes of the object x that it is F . This is also in the spirit of what David Kaplan himself argued in “Quantifying In.”¹⁹ If this is correct, then it applies also to canonical concepts of Thoughts. From

(10) a believes $\langle\text{John}\rangle^{\wedge}\langle\text{believes}\rangle^{\wedge}\langle\langle p \rangle\rangle$

it would then follow that

(11) Concerning the Thought $\langle p \rangle$: a believes John believes it.

The form of transition saliently instantiated by the step from (10) to (11) is one that will concern us further. I label it ‘exportation of concept’—since that is what it is. Exportation of concept involves the same relation between concepts of concepts and concepts one level below them as the relation that Quine labeled ‘exportation’ in the case of concepts of objects vis-à-vis the objects themselves.

We can put the validity of exportation for canonical concepts of concepts to work. It contributes to a partial explanation of the validity of the inferential pattern from (5) to (6) above in the argument for the Antithesis; and it does so in a way consistent with the soundness of the arguments for the Thesis. It contributes to an explanation of validity,

17. C. Peacocke, “Entitlement, Self-Knowledge and Conceptual Redeployment.”

18. W. Quine, *The Ways of Paradox and Other Essays* (Cambridge, Mass.: Harvard University Press, 1976), 190.

19. “Quantifying In,” sec. 8 on ‘standard names’.

because we can export on the canonical $\langle\langle\text{Bush}\rangle\rangle^{\wedge}\langle\langle\text{is powerful}\rangle\rangle$ in the premise

$\text{Bel}(\text{John}, \langle\text{Bel}\rangle^{\wedge}\langle\text{Karl}\rangle^{\wedge}\langle\langle\text{Bush}\rangle\rangle^{\wedge}\langle\langle\text{is powerful}\rangle\rangle))$.

Nonetheless, the defender of the Antithesis may reasonably insist that this is still not a full explanation of the datum he cited. For, to say it one last time, this point does not address the intuition that *no* additional premise is needed in English to explain why the transition from (5) to (6) is valid. I offer a second step and a fuller explanation, drawing on additional resources, in the next section.

Issues about partial understanding also produce interesting wrinkles here too. It is a question whether exportation is valid, even for standard names, when understanding is merely partial. Consider someone who only partially understands 'aleph-two'. He may know that it is a name (strictly, a description) of an infinite number, but may know no more than that. It may be true to say of him in English "He believes that aleph-two is larger than the number of natural numbers". But does he have relational beliefs, of aleph-two, that it has such-and-such properties? There is certainly some resistance to saying so. This resistance is increased when we reflect that what motivates relational ascriptions in some other cases of standard names of abstract objects is grasp of the notation of such a kind that the thinker knows where some number comes in an ordinal or cardinal system. Precisely that is lacking in the case of our thinker who only partially understands 'aleph-two'. There is an intuition in favor of saying that he has beliefs to the effect that aleph-two is thus-and-so; but our thinker does not know of any particular infinite number that it is thus-and-so (and in particular does not know that it is aleph-two). The Quine-like treatment of exportation is correct only if we tacitly assume a background of sufficient understanding, an assumption that goes beyond what is required for correct *de dicto* attributions. In this particular corner of the territory, however, canonical concepts of concepts and Thoughts are better-off than concepts like *aleph-two*. If grasp of the canonical concept of the concept F involves tacit knowledge of the reference rule (*), which in turn requires possession of the concept F, one cannot employ $\text{can}(\text{F})$ in thought without possessing the concept one is thinking about, and knowing which concept it is (even if one's grasp of *it* is partial). By contrast, knowing that an infinite number is two up in the aleph-ordering falls short of knowing which infinite number it is, if your grasp of the nature of the aleph-ordering is weak.

4. RELATION TO SINGLE-LEVEL TREATMENTS

In his book *Frege: Philosophy of Language*, Michael Dummett proposes a well-known alternative to the Fregean infinite hierarchy of concepts. This alternative is also considered, among other alternatives, in Terence

Parsons's discussion of these issues in "Frege's Hierarchies." On this alternative, the notion of a concept is not abandoned, but it exists at only one level. According to this treatment, in sentences such as our (2) above

(2) Karl believes that Bush is powerful

the words 'Bush' and 'is powerful', at those occurrences, not only denote their usual senses (as Frege said) but also express their normal senses (as the Fregean hierarchy denies). Under this Dummettian treatment, in

(12) John believes that Karl believes that Bush is powerful

'Bush' at its occurrence there, doubly embedded, denotes its normal sense, rather than a canonical concept of a sense. As Dummett writes about his emendation of Frege's theory, "With this emendation, there is no such thing as the indirect sense of a word: there is just its sense, which determines it to have in transparent contexts a reference distinct from its sense, and in opaque contexts a referent which coincides with its sense. There is therefore no reason to think that an expression occurring in double *oratio obliqua* has a sense or a reference different from that which it has in single *oratio obliqua*."²⁰

There are at least two different approaches to giving a formal object-language suitable for a conception on which there is only one level of sense. One approach employs a notion of denotation that is relativized to linguistic context. Here the object-language itself is nonhierarchical. Whether an expression has its normal reference or refers instead to its normal sense depends upon the sentential context in which it occurs. This style of approach is what Tyler Burge calls 'Method I' in his paper "Frege and the Hierarchy." The type of object-language it employs is one for which I argued one can give an adequate truth theory language in "Entitlement, Self-Knowledge and Conceptual Redeployment."

The other style of approach to giving a formal object-language suitable for a conception on which there is only one level of sense is to introduce terms—simple or complex, variables or constants—that correspond in the formal language to expressions of English consisting of 'that...' clauses. This is Burge's 'Method II' in "Frege and the Hierarchy." Under this treatment, denotation does not for these purposes need to be treated as relative to linguistic context. In a formal language that makes explicit to which concepts and Thoughts a subject stands in various relations, under this second form of single-level treatment, (12) could be regimented thus:

(12S) Believes (John, <Karl>^<Believes>^<Bush>^<is powerful>).

We can call the single-level language in which (12S) is cast 'SL'.

20. Dummett, *Frege: Philosophy of Language*, 268.

By contrast with both of these styles of approach, under the treatment with the Fregean hierarchy explicitly articulated in the object-language, as we said, (10) would be regimented

(12H) Bel (John, < Bel>^ <Karl>^ <<Bush>>^<<is powerful>>).

Henceforth I distinguish 'Bel' and 'Believes' in order not to prejudge any issue of the identity of the respective belief-relations used by the different kinds of treatment. I reserve 'Bel' for the explicitly hierarchical language.

Is there a translation scheme between formulae of the single-level language SL of (12S) and the hierarchical language of (12H)? There is; in fact, there is more than one. I distinguish a superficial translation scheme and a deeper translation scheme.

The superficial translation scheme is the one that is simply obvious from the notation itself, independently of any substantive theses about concepts. In any sentence like (12S), we can add to expressions like "<is powerful>" additional pairs of pointed brackets, according to this rule: We add one additional pair for each embedding of the occurrence of the expression within the versions of expressions that are traditionally regarded as creating oblique contexts. Unembedded occurrences are left unchanged. If we then also change 'Believes' to 'Bel', the result is a translation from a sentence of the single-level language to a sentence of the hierarchical language. Applying these rules to (12S) yields (12H) as its translation.

The significance of this translation lies in the fact that the defender of the hierarchy could, if he so wished, use the same sentences as this form of single-level theorist, but regard these sentences as legitimate only insofar as they have a translation into his hierarchical language. The friend of the hierarchy could insist that the translation into the hierarchical language makes perspicuous and explicit the structure that exists in the realm of concepts and Thoughts.

This is, however, a superficial translation scheme from the point of view of answering the 'argument from inferences', because the scheme does not in itself, independently of any further theses, involve any commitment to the idea that when (12S) is true, John is really related by a belief relation only to the Thought

<Bush>^<is powerful>

itself, as opposed to a concept of that Thought.

If the principle of exportation of concept of the preceding section is correct, however, there is also a deeper translation scheme available for the single-level language SL. This deeper scheme does imply that in the translation of (12S), a belief relation is represented as holding between the subject and the Thought

<Bush>^<is powerful>

itself. Exportation of concept implies that when someone has an attitude to a content containing a canonical concept of a Thought, he has the corresponding *de re*, relational attitude to that Thought itself. This applies to the concept

$\langle\langle\text{Bush}\rangle\rangle^{\wedge}\langle\text{is powerful}\rangle\rangle$

as it is mentioned in (12H). In the presence of these claims, (12H) implies (12HR):

(12HR) Concerning the Thought $\langle p \rangle$, viz. $\langle\text{Bush}\rangle^{\wedge}\langle\text{is powerful}\rangle$,
Bel (John, $\langle\text{Bel}\rangle^{\wedge}\langle\text{Karl}\rangle^{\wedge}\langle p \rangle$).

Under the deeper translation scheme, (12S) is translated as (12HR). (12HR) manifestly involves a relation to the Thought $\langle p \rangle$ itself. The general rule for the deeper translation scheme can be given in two steps. First, we take a sentence of SL and translate it into a sentence of the explicitly hierarchical language in accordance with the previously mentioned, more superficial, translation scheme. We then export on the canonical concepts corresponding the most deeply embedded 'that...' clauses. (We could of course also validly export on intermediate 'that...' clauses; for some inferences, it will be important to do that.)

If we use the deeper translation scheme, we can speak with the user of the single-level language SL. The Dummettian language is undeniably more concise, less cumbersome, than the explicitly hierarchical language. Under the deeper translation scheme, the language is not even misleading in representing the truth of (12S) as involving a relation to the embedded Thought $\langle p \rangle$. We can speak this way without denying the fundamental arguments for the hierarchy, the arguments for the Thesis.

The deeper translation scheme finally puts us in a position to take a second step in answering the 'argument from inferences' that featured in the case for the Antithesis. That argument appealed to the validity of the transition from

(5) John believes that Karl believes that Bush is powerful

to

(6) There is some proposition $\langle p \rangle$ such that John believes that Karl believes $\langle p \rangle$, and $\langle p \rangle$ is the proposition that Bush is powerful.

(5) is regimented in SL as (12S), and (12S) in turn is, under the deeper scheme, translated as (12HR). But (12HR) immediately logically implies (6), as required. It does so, crucially, without relying on any additional premises special to the subject matter. The transition from (12HR) to (6) relies only on first-order logic with identity.

Similar considerations apply to the other illustration mentioned in the 'argument from inferences', that from (5) together with

(7) It is true that Bush is powerful

it follows that

(8) Something John believes Karl believes is true.

Under the deeper translation scheme, (5), (7), and (8) are translated respectively as (12HR), (13HR), and (14HR):

(12HR) Concerning the Thought $\langle p \rangle$, viz. $\langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$,
Bel (John, $\langle \text{Bel} \rangle^{\wedge} \langle \text{Karl} \rangle^{\wedge} \langle p \rangle$)

(13HR) Concerning the Thought $\langle p \rangle$, viz. $\langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$,
True($\langle p \rangle$)

(14HR) There is some Thought $\langle q \rangle$ such that
Bel (John, $\langle \text{Bel} \rangle^{\wedge} \langle \text{Karl} \rangle^{\wedge} \langle q \rangle$) and True ($\langle q \rangle$).

(14HR) follows from (12HR) and (13HR), again without any additional premises special to the theory of concepts, just as the ‘argument from inferences’ demanded. This explains away the last consideration offered in support of the Antithesis, consistently with acceptance of the philosophical case for the hierarchy.

Experts on the regimentation of propositional attitudes will have realized for some pages that I have been suppressing a complication. When Karl has the belief that Bush is powerful, Karl stands in a propositional attitude relation to the Thought $\langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$ itself, and not to a Thought given under a mode of presentation. When John has a belief about Karl’s belief, matters stand differently. When John believes that Karl believes that Bush is powerful, John is thinking in a quite particular way of the Thought to which Karl stands in the belief-relation—he is thinking of it under its canonical concept. The English sentence “John believes that Karl believes that Bush is powerful”—our (5)—is most naturally read as attributing to John use of this way of thinking of the Thought to which Karl stands in the belief relation. But our regimentation (12HR) does not in itself imply that John is thinking of the relevant Thought in this way.

(12HR) and a certain background assumption do jointly imply that John is thinking of the Thought that way. The Background Assumption is that a thinker can have a *de re* attitude about a Thought only by thinking about it under its canonical concept. Under the theses of this chapter, this in turn implies that a thinker can have *de re* attitudes about a Thought only if he is capable judging and having other attitudes to that Thought. The Background Assumption seems to me to be true. It is hard to conceive of what it would be to have a *de re* attitude to a Thought without having the ability to think that Thought. Intuitively, it seems that if someone were not able to think the Thought in question, he would be able to have only *de dicto* attitudes concerning the Thought. If the Background Assumption is true, there is no gap between the truth of (12HR) and the truth of (5). If the Background Assumption is correct, it too iterates up the hierarchy of concepts. To have a *de re* attitude about the canonical concept of a Thought is to think of that canonical concept

under its canonical concept; and so forth. For those who like symmetry in their theories, we can note that if the Background Assumption is true, we have for canonical concepts not only the exportation transitions we discussed earlier, which export a term outside a propositional attitude verb, but also a kind of importation inference. When a thinker has a relational, *de re* attitude to a Thought, we can pull to the inside of the scope of the relevant propositional attitude verb the term canonically specifying that Thought.

But suppose we wanted to give an account of the English sentences that does not presuppose the truth of the Background Assumption—how should we then proceed? I will not go into great formal detail at this stage of this essay, but I will mention one line of development to illustrate further possibilities. In his paper “Reference and Propositional Attitudes,” Brian Loar argued some years ago now that some terms within the scope of propositional-attitude verbs make what he calls ‘a dual contribution’ to the truth-conditions of the whole in which they occur.²¹ In some cases, it is important to recognize that a term both contributes a concept or mode of presentation to the intentional content of the attitude, and also indicates that the subject of the attitudes stands in some relational attitude to the reference of the same term. Loar’s point could be applied to canonical concepts of Thoughts and concepts. In the translations like (12HR) through (14HR) from the single-level language SL, we have taken only the reference of the canonical concepts of Thoughts as relevant to the truth-conditions. But in considering the issue of translations back to sentences of SL, there are advantages in regarding the canonical concepts as making exactly the kind of dual contribution that Loar describes. They both specify how a thinker is thinking of a Thought—precisely what we were discussing in the previous two paragraphs—and, by their reference, which Thought it is he is ascribing to the subject whose attitudes he is thinking about. So on a more complex implementation of the deeper scheme, our translations would not merely be sentences like (12HR) through (14HR). In the more complex scheme, they would include those conditions but would also add explicitly how, for instance, John is thinking of the Thoughts he is attributing. Instead of (12HR), we would have something that says:

Concerning the Thought $\langle p \rangle$, viz. $\langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$, John believes of it, under the concept $\text{can}(\langle p \rangle)$, that it has the property that Karl stands in the belief relation to it.

This can be formalized straightforwardly. Corresponding formalizations will still validate the transition from (5) and (7) to (8).

However we implement the deeper translation scheme, either in a simple or in this more complex fashion, the single-level language SL used

21. B. Loar, “Reference and Propositional Attitudes,” *Philosophical Review* 81 (1972): 43–62.

in formulae like (12S) raises semantical and formal issues that need to be addressed. I make a start on these in the appendix to this chapter. SL is closer in some respects to natural language, and this makes it attractive as a means of describing embedded attitudes. But this closeness and attractiveness should not be taken as evidence in support of a single level of concepts. The very existence of the deeper translation scheme that legitimizes use of SL undermines any idea that the single-level language is intelligible only if there is a single level of concepts. Hierarchies of sense and understanding need not be reflected in hierarchies of expressions in our natural language. The theory of actual linguistic structures is one thing. The philosophical theory of concepts and understanding is another.

5. AN APPLICATION: RECONCILING EXTERNALISM WITH DISTINCTIVE SELF-KNOWLEDGE

One application of the Leverage Account is in answering arguments that externalist treatments of intentional content cannot be reconciled with a subject's having a distinctive kind of knowledge of his own mental states. Twenty years have now passed since intensive discussion began about the reconcilability of distinctive self-knowledge with the external character of intentional content. The first decade of such discussions has by no means quieted the objections of those who say that some theories of self-knowledge make such reconciliation impossible, so that we must either abandon those theories or reject externalism about content. The questions arise sharply, for instance, for a theory that states that we have a species of action-awareness of our own mental actions, and which states that this is what makes available our knowledge of our own mental actions.²² Is that an account that makes such reconciliation impossible? And if it does not, what positive account is to be given of the nature of the reconciliation?

Specific doubts about the possibility of reconciliation on certain models of self-knowledge have been concisely articulated, and endorsed, by Crispin Wright, writing about halfway (1996) through this twenty-year period. Wright considered the model of self-knowledge as inner observation, and wrote: "I want to say that... in the sense in which an image or mental picture can come before the mind, its intentionality cannot."²³ "Both a sunburned arm and a triangle can be presented as ordinary objects of observation, and each sustains, *qua* presented under those particular respective concepts, certain internal relations: the sunburned arm to the causes of its being in that condition, and the triangle

22. See my essay "Mental Action and Self-Awareness (I)," in *Contemporary Debates in the Philosophy of Mind*, ed. J. Cohen and B. McLaughlin (Oxford: Blackwell, 2007), 358–376.

23. C. Wright, *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations* (Cambridge, Mass.: Harvard University Press, 2001), 342.

to, for instance, other particular triangles. And the point is simply that while the identification of the triangle as such can proceed in innocence of its internal relations of the latter kind . . . recognition of the sunburned arm as just that cannot proceed in like innocence but demands knowledge that its actual causation is as is appropriate to that mode of presentation of it."²⁴ Wright attributes to Wittgenstein, and finds convincing, the point that "the internal relations to the outer, of whatever sort, are all of the latter—sunburn-style—kind; and hence there is indeed a standing puzzle in the idea that an appropriate characterization of them, incorporating such intentionality, is somehow vouchsafed to their subject by something akin to pure observation."²⁵

Does Wright's objection apply equally to the action-awareness account? In the perceptual case, as Wright implied, there is a partition between properties such as that of being sunburned that cannot be known to be instantiated simply by taking perception at face value, and those such as shape, color, orientation, surface texture, and so forth, that can be known to be instantiated simply by taking perceptual experience at face value. (A state's representing these latter properties is also a matter of its external relations—a fact, incidentally, that should give us pause about the direction in which the argument is going.) An analogous partition of properties, as thought about in given ways, can be made for action-awareness. I cannot, from action-awareness alone, come to know that the copying machine whose lid I am closing was manufactured in Taiwan. Knowing that requires knowledge of its history that is not given in action-awareness itself. But action-awareness can make available knowledge that I am closing the machine's lid, at a certain speed, with a certain force, and that I am doing it now. So, in the case of mental actions, the crucial question to address is this: Is the intentional content of a mental event or state to be grouped with the property of being made in Taiwan, or is it to be grouped with the properties that you can know about simply from your action-awareness?

The intuitive, pretheoretical answer to this question is that we have an action-awareness of the full intentional content of our judgments, decisions, and other mental actions. We are aware that we are judging that New York is hot in the summer; we are aware that we are deciding to spend the summer in a cooler place. A judgment with that intentional content may also be a manifestation of a neurosis, may be an unconscious excuse for not staying in New York, or many other things that are to be grouped with the machine's being made in Taiwan. But in ordinary cases, the intentional content of the judgment, decision, or whatever mental action-type is in question does seem to be so available. What is the explanation of this fact?

24. *Rails to Infinity*, 343.

25. *Rails to Infinity*, 343.

Whatever the explanation, it will have to have a certain generality. When we know what we are judging or deciding, on the basis of action-awareness, we know the content of our judgment or decision, whatever its conceptual constituents. You can have an action-awareness of your judging that p , whatever the content p may be, whether the conceptual content p is observational, theoretical, moral, or anything else. As we noted earlier, in such judgments about the content of mental states, you think of the conceptual constituents of the content judged under their canonical concepts. You think of the concept *hot* under its canonical concept; and so forth.

We can now focus on the transition a thinker makes in passing from an action-awareness of his judging that New York is hot in the summer to his judgment of the following intentional content, where the concepts referred to are thought about under their canonical concepts:

$\langle \text{Judge} \rangle^{\wedge} \langle \text{I} \rangle^{\wedge} (\text{can}(\langle \text{hot} \rangle)^{\wedge} \text{can}(\langle \text{New York} \rangle)^{\wedge} \text{can}(\langle \text{in the summer} \rangle))$.

(Purists can add notation for the canonical concept of $\wedge$, predicational combination in Thoughts, but I will abstain in the interests of legibility.) This transition from the action-awareness to the judgment is a priori valid. In any context in which the thinker has a genuine, and not merely apparent, action-awareness of judging that New York is hot in the summer, it will also be true that he judges that it is the concept *hot* that he judges New York to fall under in the summer. The same applies to the canonical concept of any other concept F in place of the concept *hot*, however externally or historically individuated the concept F may be. This explanation has the required generality.

This transition in thought from action-awareness to judgment is totally different in character from the transition, unwarranted without further information, from

a perception of a reddish arm

to a judgment

this arm is sunburned.

This latter transition does, just as Wright said, need further information about the causes of the redness on the arm if it is to be valid. By contrast, the preceding transition involving action-awareness does not need any further information for its legitimacy. No further information is needed, under the Leverage Account, because if the conceptual constituents of the content of the judgment of which there is action-awareness are externally individuated, so are the canonical concepts of those constituents. For the transition involving action-awareness, the reason-giving state and the judgment for which it gives reason are equally externally individuated. There is not even an apparent leap from what is internally individuated to what is externally individuated.

There may be a sense of unease about this reconciliation of externalism and the distinctive knowledge of mental actions, a sense that there is some kind of cheating going on. This unease may stem from the thought that the treatment given here is like that of someone who insists, correctly, that the recognitional concept of water is externally individuated, and that we know our thoughts are water-thoughts. There is a clear sense in which one can possess a recognitional concept of water without knowing which liquid it is, in the sense of not knowing its chemical composition. Does a similar objection apply against the account I have offered of action-awareness of the conceptual contents one is judging? Is the account consistent with the thinker's not knowing which concepts are in question?

I reply that because one is employing the canonical concept of a concept in making judgments about the contents of one's thoughts, one does, by contrast with the chemical characterization of water, know which concept is in question. It is precisely the force of the Leverage Account to make it clear that any such seeming gap is really closed. Under the Leverage Account, you know as much about which concept is in question when you think of it as *the concept F* as there is to know. (You may learn more by philosophical investigation about the nature of the concept, but that is not a matter of further determination of which concept is in question. You already knew that in fixing on just one concept to investigate.) All the conditions that contribute to the individuation of the concept *F* itself contribute to the individuation of the canonical concept of *F*. From the Leverage Account, as noted, reasons for making first-order judgments containing the concept *F* are transmitted to reasons for making suitably corresponding judgments containing the higher-order concept *the concept F*. Any requirements on knowing which concept that are met when one is simply using the concept *F* will equally be met when one uses the canonical concept of the concept *F*. It is always an answer to the query "Which concept is in question?" to say "It's the concept *F*," where this answer employs the canonical concept of *F*, rather than some descriptive mode such as "the concept discussed in chapter 5 of such-and-such book."

There will, for any given concept, be empirical conditions met by a given thinker who employs that concept, conditions not extractable simply from the nature of the concept itself. They will include such matters as the nature of that particular thinker's mental representations underlying his possession of the concept, and the particular computational procedures involving it that he employs. There is manifestly an important area of study that consists in the empirical investigation of these empirical matters involving concepts as possessed by particular thinkers. But precisely because these empirical conditions can vary across thinkers that share the same concept, these empirical conditions are not what constitute the nature of the concept itself. A thinker's ignorance of these empirical matters of mental representation does not impugn her

knowledge of which concepts are in question when she thinks, "I judge that New York is hot in the summer."

It is a striking feature of the canonical concept of a concept that it has two characteristics whose coinstantiation rests on a merely empirical truth. The canonical concept has the individuating properties specified in the Leverage Account. Our minds and conscious states are also such that we can rationally apply the canonical concept of a concept in response to conscious states, such as action-awareness and passive thinking, whose content involves the very concept of which it is a canonical concept. It seems to be a precondition of rational, critical thought that these two characteristics go together. We are sometimes required rationally to consider, for instance, whether our conscious judgment that Fa was made in an epistemically responsible fashion. Investigation of this issue involves drawing on our tacit knowledge (or tacit partial knowledge) of the condition for something to be F. Such a rational exercise of thought is possible only because the canonical concept of F is one we can apply in rational response to conscious mental states whose intentional content contains the concept F.

Canonical concepts of concepts are far from the only concepts some of whose distinctive applications rest on empirical facts. The way we think of a type of bodily movement, when we perceive it made by someone else, yet also perceive it as an action of a type that we ourselves could make, provides another type of example. No doubt the underlying ground of the possibility of such concepts involves the now famous 'mirror neurons' identified by Rizzolati and his colleagues.²⁶ It is an empirical matter that there are such representations in our psychology. They make possible much that would not otherwise be possible. To deny the existence of ways of coming to apply concepts that rely on empirical facts would rule out large tracts of human thought and experience. This applies equally to our ability to know about the intentional content of our own mental actions and our other conscious states.

APPENDIX: SEMANTICAL AND FORMAL ISSUES

The main text of this chapter has concentrated on core philosophical issues about concepts of concepts, and their relation to understanding. So as not to interrupt the flow of that discussion, I have collected some more technical issues in this appendix. The technical issues are nonetheless important, and sometimes pivotal. Some of the stances taken in the recent literature would, if correct, undermine various features of the position for which I have been arguing. Other stances on formal or semantical issues would rule out the single-level language to which

26. G. Rizzolati et al., "Premotor Cortex and the Recognition of Motor Actions," *Cognitive Brain Research* 3 (1996): 131–141.

I have said that a defender of the Thesis can have access. These formal and semantical issues may be technical, but they are not technicalities.

I start with the most pressing questions for the position I have developed here and then open the discussion into a consideration of the relation of this position to some others in the literature.

Question 1: Does a hierarchy of canonical concepts collapse into, or at least make available, a single-level account of sense and concepts?

In his early paper "Frege's Hierarchies," Terence Parsons argues that the hierarchy does so collapse. He classifies as 'rigid' any theory that holds that "the customary sense of an expression uniquely determines its indirect sense" (44). The theory I have offered is certainly rigid in this sense. In Parsons's notation, $s_1[A]$ is the customary sense of A ; $s_2[A]$ is the unique indirect sense of A , what I have called the canonical sense of $s_1[A]$; and, where a and b are senses, $a(b)$ is the result of "applying" the sense a to the sense b . (This last embodies a controversial conception of sense and/or controversial interpretation of Frege, but we can equally construe $a(b)$ as the predicational combination of sense a with sense b for purposes of the present discussion.) Parsons goes on to argue:

A rigid theory can be converted into a theory that associates with each expression only one sense, as follows. We define "*the* sense of A ", i.e., " $s[A]$ ", as follows:

- (i) If A is a name or a sentence or a predicate or an extensional sentence operator, then $s[A] = s_1[A]$.
- (ii) If A is an indirect sentence operator, then $s[A] =$ that function which maps an arbitrary sense x to $s_1[A](s_2[B])$, where B is an expression such that $s_1[B] = x$. (45).

I do not dispute that we can define the operation Parsons specifies, but I do not think that that shows the hierarchy is eliminable or redundant. The complex sense $s_1[A](s_2[B])$, in terms of which '*the* sense of A ' is individuated, involves a second-level sense and could not be grasped by a thinker unless he grasps the canonical sense of a first-level sense. It seems to me that we do not have a reduction of senses to a single level if an account of what it is to grasp senses in that single level involves grasp of senses at other levels. To insist that that is irrelevant to reduction is to set sense loose from its essential connections with grasp of sense or understanding, by intentional agents and thinkers. This is of course not a criticism of Parsons if his intentions either were purely definitional or were concerned with a kind of ontological reduction that is not constrained by considerations of understanding. Parsons does, however, (in passing) write that a variant of this approach just described turns out to be the theory "that Dummett says Frege should have given" (45). Dummett certainly was intending to offer an alternative to the hierarchy.

Question 2: Do representations in the single-level language SD together with obvious truths imply falsehoods?

As we noted in the main text, under the treatment Tyler Burge called "Method II," a language designed originally for only a single level of sense formally uses terms in the position corresponding to the complete sentences of English that are embedded in propositional attitude operators. In his paper "Frege and the Hierarchy," Burge argued that, in the presence of other Fregean principles, this treatment implies that if someone believes one truth, he believes every truth (pp. 271–272 in the original paper). If sound, this is effectively a *reductio* of Method II when developed without a hierarchical language, in the presence of the other Fregean principles.

I accept that Burge's argument is valid given his premises and his interpretation of the language. I also accept, though for other reasons, the conclusion that Burge draws from this argument (and others) that there is good reason to introduce a hierarchy of concepts. The pressing question that arises for my position is this: Why do representations in the language SL, such as my (12S)

(12S) Believes (John, $\langle \text{Karl} \rangle^{\wedge} \langle \text{Believes} \rangle^{\wedge} \langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$)

not fall victim equally to Burge's *reductio*? For in the formulae of SL, the English sentences embedded in propositional attitude contexts are certainly represented by terms. The term

$\langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$

is one such; the term

$\langle \text{Karl} \rangle^{\wedge} \langle \text{Believes} \rangle^{\wedge} \langle \text{Bush} \rangle^{\wedge} \langle \text{is powerful} \rangle$

is another.

I am going to quote the initial statement of Burge's argument in full. (I change the numbering of examples to avoid confusion with numbered examples already used; the altered numbering is indicated by square brackets.) Burge started by considering an example of double embedding of the sort we have been considering:

(15) Igor believes Bela believes Opus 132 is a masterpiece.

He continued:

Let us assume then that ' α ' (which denotes the proposition that Opus 132 is a masterpiece) represents 'Opus 132 is a masterpiece' as it occurs in [15]. [15], we shall assume, asserts a relation of belief between Igor and the proposition that Bela believes that Opus 132 is a masterpiece. I shall denote this proposition by the expression ' $\Gamma_1(\beta_1, \alpha)$ '. Thus on our assumptions, [15] is formalized as

[16] Believes (Igor, $\Gamma_1(\beta_1, \alpha)$)

By the principle of extensionality, the denotation of ' $\Gamma_1(\beta_1, \alpha)$ ' is a function of the denotations or extension of its parts. I shall assume that ' β_1 ' denotes

the sense of 'Bela' and that ' T_1 ' denotes the sense of 'believes'—a function from β_1 and α to the relevant proposition.

We assume the principle that a given sense is associated with a unique denotation or extension. Thus the proposition $\Gamma_1(\beta_1, \alpha)$ is associated with (or, in Church's terminology, is a concept of) a unique denotation or extension, its truth value.

We assume that this truth value is a function of the unique denotations or extensions associated respectively with the senses that determine the proposition. . . . Let ' β ' express β_1 and denote Bela; let ' T ' express Γ_1 and denote what 'believes' denotes (or have its extension). Let α_0 express α and denote its truth value. (We suppose that truth value to be truth.) Then ' $T(\beta, \alpha_0)$ ' expresses $\Gamma_1(\beta_1, \alpha)$ and denotes its truth value. 'Believes' originally applied to persons and propositions. But on our assumptions it has come also to apply to persons and truth values. This leads to absurdity in short order.

For given the classical substitution laws of Method II, we may substitute any expression that denotes truth for ' α_0 ' in ' $T(\beta, \alpha_0)$ ' and preserve the truth value of ' $T(\beta, \alpha_0)$ '. (We speak of both sentences and propositions as having truth value.) But ' $T(\beta, \alpha_0)$ ' supposedly expresses the proposition that Bela believes Opus 132 is a masterpiece. So it seems to follow that if Bela believes Opus 132 is a masterpiece, he believes every truth.

The argument shows that on these assumptions 'Opus 132 is a masterpiece' in [15] cannot be represented by a term ' α ' denoting the proposition that Opus 132 is a masterpiece. (271–272)

Representations such as (12S) in the single-level language SL are not vulnerable to this argument because one of the assumptions of the argument fails for SL. Terms of SL such as '<Bush>^<is powerful>', '<Opus 132>^<is a masterpiece>' denote Thoughts, or complex concepts. Those Thoughts or complex concepts do also have truth-values, but those truth-values *are not the denotations or semantic values of these terms in the language SD*. SL is nevertheless still an extensional language. The denotation of the complex term '<Opus 132>^<is a masterpiece>', a complex Thought, is a function of the denotation of its constituent terms '<Opus 132>' and '<is a masterpiece>' (and of the denotation of the functor " $\wedge$ "). There is no difficulty on this view in the formula

(17) Believes (Bela, <Opus 132>^<is a masterpiece>)

being true, while the formula

(18) Believes (Bela, <The Continuum Hypothesis>^<is independent of>^<the axioms of ZF>)

is false. The Thought denoted by the term '<Opus 132>^<is a masterpiece>' is distinct from the Thought denoted by the term '<The Continuum Hypothesis>^<is independent of>^<the axioms of ZF>'.

Can it be objected to this argument that we can still make the substitution of any true sentence for 'Opus 132 is a masterpiece' in the English

sentence 'Bela believes Opus 132 is a masterpiece'? That would be begging the question against SL in the absence of the assumption that the terms that correspond to complete sentences denote truth-values. The hypothesis is that formulae of SL, like (17) and (18), translate the corresponding English sentences; and under that hypothesis, substitution of sentences with the same truth-value within 'believes that' contexts is clearly not always going to preserve truth-value. The translations of (17) and (18) into the explicitly hierarchical language, under the deeper scheme, simply underwrite this point.

Does the rejection of the assumption on which the *reductio* rests make the solution offered by SL un-Fregean? The complete sentences of SL, as opposed to terms for Thoughts, can and should be regarded as denoting truth-values, in classical Fregean fashion. And the fully hierarchical language into which I have suggested some deeper translation schemes for sentences of SL seems to me wholly Fregean. What I have resisted is the substitution of a complete sentence such as "Opus 132 is a masterpiece" (as opposed to a term) in the position of '<Opus 132>^<is a masterpiece>' in such SL-formulae as (17). In the formulae subject to Burge's *reductio*, unlike SL, complete sentences can occupy the same positions as terms for Thoughts. (This is legitimate under his assumptions about denotation and semantic value.) We could if we so wished design an extension of the language SL, a language SL+, one that does allow such substitutions of a sentence in certain term-positions of formulae of the unextended language SL as well-formed. But the reference (semantic value) of the sentence thus allowed in term position would have to be its normal sense. This reference-shift too would also seem to be entirely Fregean in spirit.

CHAPTER 11

Directing Intentions

John Perry

1. INTRODUCTION

In his monograph *Demonstratives*,¹ David Kaplan distinguishes between pure indexicals such as “I,” “here,” and “now” and true demonstratives such as “this,” “that,” “he,” and “she” (in some of their uses). The whole set of context-sensitive expressions are officially called “indexicals;” the title of the monograph was chosen for historical reasons. I will use “demonstrative” here to mean true demonstratives and “indexicals” to mean pure indexicals.

Kaplan gives us four theories of demonstratives, three in *Demonstratives* and a fourth in “Afterthoughts.” In *Demonstratives* we find the Fregean Theory, which Kaplan thinks is wrong, the Indexical Theory, which he thinks is not so much wrong as a “mean thing” (528), and the Corrected Fregean Theory, which he advocates there. In “Afterthoughts,”² he replaces the Corrected Fregean Theory with the Directing Intentions Theory, or at least is strongly inclined to do so. (I don’t think he officially gives his final theory a name, but that is the name I will use.)

It’s clear from this that Kaplan finds demonstratives puzzling. He offers only one theory of indexicals; once the right distinctions are made, the facts about the semantics of indexicals are “obvious and incontrovertible.” Clearly the case is otherwise with demonstratives. He says at the end of his discussion in “Afterthoughts”: “There is something I’m not understanding here, and it may be something very fundamental about the subject matter of logic” (590).

1. David Kaplan, *Demonstratives*, in Joseph Almog, John Perry, and Howard Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), 481–564; subsequent references to “*Demonstratives*” are given in parentheses.

2. David Kaplan, “Afterthoughts,” in Joseph Almog, John Perry, and Howard Wettstein, eds., *Themes from Kaplan* (New York: Oxford University Press, 1989), 565–614; subsequent references to “Afterthoughts” are given in parentheses.

I will develop what I take to be version of Kaplan's Directing Intentions Theory and argue in its favor, and indicate how this version might help with the logical problems that vex Kaplan.

2. FOUR THEORIES OF DEMONSTRATIVES

According to Kaplan, the character of an expression is a function from contexts to appropriate contents. Proper contexts are quadruples of agents, times, locations, and worlds such that the agent is at the location at the time in the world, that is, c_A is at c_L at c_T in c_W . The character of "I," for example, is function that delivers, at each context, the agent of the context as the referent of "I." The characters of words are determined by the conventions of language.

The Indexical Theory, the Corrected Fregean Theory, and the Directing Intentions Theory all apply the Context-Character-Content (or CCC) structure to demonstratives. It remains incontrovertible that demonstratives are devices of direct reference. The problem is that the CCC structure does not seem to give us enough to get us from the occurrence of a demonstrative to a referent. Consider the context consisting of me, now, here, in this world. There are dozens of things I could refer to with "that," without abusing English: the essay I'm writing, the computer I'm writing it with, the irritating fleck on my glasses, the can of diet Sprite on the table beside me, and so forth. Fixing the conventions of English and the context aren't enough. Something else is needed.

Once we have found the additional factor that we need, there is another decision to make. Where in the CCC theory to put it: expression, character, context, or content? Table 11.1 summarizes the differences between the four theories on these issues:

Table 11.1.

Theory	What Fills the Gap	Where It Goes in the Theory
The Fregean Theory	A demonstration, thought of as a relation δ , which determines a demonstratum at a time and location	In the proposition expressed
The Indexical Theory	The demonstratum itself	In the context, as an additional parameter
The Corrected Fregean Theory	Same as the Fregean Theory	In the character of the augmented demonstrative itself; "That" becomes "That[δ]"
The Directing Intentions Theory	A directing intention δ , to refer to the object of a perception	Same as above

3. DEMONSTRATIONS, INTENTIONS, AND WHAT IS SAID

As I said, in “Afterthoughts” Kaplan is at least inclined to replace the Corrected Fregean Theory with the Directing Intentions Theory. In *Demonstratives* he contrasts the determinative role of demonstrations in the case of demonstratives, to the case in which a person points at himself while saying “I.” In the latter case the gesture is not criterial but simply a sort of helpful add-on. In “Afterthoughts” he demotes demonstrations to playing this helpful add-on role, while directing intentions become criterial.

If we trace back the conceptual apparatus of the CCC, we see that content is introduced in terms of the more or less pretheoretical concept of “what is said.” For example, here is how one of Kaplan’s crucial arguments begins:

I return now to the argument that all indexicals are directly referential.
Suppose I point at Paul and say,

He now lives in Princeton, New Jersey.

Call *what I said*—i.e., the content of my utterance, the proposition expressed—‘Pat’. Is Pat true or false? True! Suppose that unbeknownst to me, Paul had moved to Santa Monica last week. Would Pat have then been true or false? False! (512)

Here the commonsense term “what is said” is equated with the more philosophical expression “the proposition expressed,” and intuitions about the application of the concepts of “saying the same thing” and “expressing the same proposition” are used to motivate central theses about content. Given that, the question is something like this. If someone has an intention to refer to X, but gestures toward Y instead, does he end up referring to, and saying something about, and expressing a proposition about X? or Y? Which is determinative, the intention or the demonstration?

In Kaplan’s “Dthat,”³ there is an example that is relevant to this question, which he doesn’t discuss in either *Demonstratives* or “Afterthoughts.” I’d like to discuss it for a bit. This is the case of the lecturer—I’ll call him Professor Z—who intends to point to a picture of Carnap, which he thinks hangs behind him in plain view of the audience, but instead points to a picture of Spiro Agnew, which a prankster has hung there in its place. He says, “The man in *that* picture is the greatest philosopher of the twentieth century.” Has he referred to the picture of Agnew and said something absurd? Or has he referred to the picture he intended to refer to, the picture of Carnap, and said something (arguably) true? To keep things simple I’ll shift to a slightly simpler (though less funny) version. Call the pictures C and A. Professor Z thinks he is pointing at C, although he is really pointing to A. He says, “That is a picture of Carnap.” I’ll call this Case 1. At first glance, this would seem a good test case. If we think the lecturer has referred to A and said

3. David Kaplan, “Dthat,” in P. Cole, ed., *Syntax and Semantics*, vol. 9 (New York: Academic Press, 1978).

something false, we should favor demonstrations and the Corrected Fregean Theory. If we think he has referred to C and said something true, we should favor the Directing Intentions Theory.

But at second glance, this isn't quite right. Although the lecturer has an intention to refer to C, I don't think he has a *directing intention* to refer to C. To make this point, let me introduce Case 2. The picture now hangs on a side wall of the lecture hall. The lecturer, who is myopic, can see the picture, although he cannot make out any details. He simply assumes it is C, the picture of Carnap, which has been hanging there for years. He points to it and says, "That is a picture of Carnap." This alteration brings the example more squarely into the set of thing Kaplan is inclined to say in "Afterthoughts," which are limited to perceptual demonstratives.

Now in this case it seems clear that the lecturer's directing intention is to refer to the picture *that he sees*, and that he knows the audience can see. He intends, or we might even say *plans*, to refer to C *by* referring to the picture he can see, because he believes that the picture he sees is C. The directing intention is the intention to refer to the picture he sees, not the intention to refer to C. At least that's how I interpret Kaplan. This interpretation may be accurate; at any rate, I think it is the one that makes it his view of directing intentions most plausible.

Now if we extend this understanding of directing intentions to Case 1, in which the lecturer points behind him, to something not in his own visual field, it seems that we no longer have a good test case. The directing intention is to refer to the picture behind him, hanging on the wall. He intends *by* doing that to refer to C, but referring to C is not his directing intention. On this understanding, both the Directing Intention theory and the Corrected Fregean Theory give the same result; he has referred to A, and so said something false.

Here is Case 3. Both A and C hang on the wall on the side of the lecture hall, so that both the lecturer and the audience can see them. Professor Z is not myopic. The pictures are not very close together. Professor Z realizes that C is the one closer to him, and he has a directing intention to refer to that picture, the one closer to him to which he attends. Perhaps the Agnew picture is only at the periphery of his visual field, or not in his visual field at all. But our lecturer is a careless person, and one with a limited sense of how his own body works, and he doesn't know his audience includes many philosophically challenged alumni back for homecoming, who have stepped into the lecture hall mainly to get out of the rain. It doesn't occur to him that anyone would think the person pictured in A could be any philosopher, much less Carnap. He extends his arm and finger in a gesture that any uninformed observer would take to be a demonstration of the picture hanging farther away, the picture of Agnew. So here the directing intention is to C, the demonstration to A; a real test case. What did he say?

The example seems to provide a test case in the sense that the Corrected Fregean view and the Directing Intentions view give different

results. But unfortunately, it isn't very clear which result is right. What is clear is that the lecturer intended to say something about C, and that many in the audience, through no fault of their own, took him to have said something about A. We are pretty clear about what he intended to say, and what he was understood as saying. But what did he say?

Case 3 doesn't answer our question, but it does lead us to an important point. I think we have a fairly robust concept of *what someone said*, and *what was said* by an utterance, and *when X and Y have said the same thing*. The concept is robust enough to support many of the arguments for which philosophers like Kaplan, Kripke, Donnellan, and Wettstein have used it. But in the final analysis it does not seem to be quite the right concept on which to erect semantic theory.

The problem is that there is a forensic element to our ordinary concept of what is said. Saying something is often a social act, which has effects on others in virtue of the words used, their meanings, and other publicly observable indications of the speaker's intentions ("perlocutionary effects," in Austin's terminology). One is held responsible for some of these effects. Suppose a member of the audience goes off and buys a portrait of Agnew, thinking it to be a portrait of Carnap, and hangs it in his hallway with the words "Rudolf Carnap: The greatest philosopher of the twentieth century" under it. Then he is made fun of and humiliated by his better educated friends. He could justifiably complain to the lecturer, "You *said* that [pointing to A] was a picture of Carnap, whom you also said was the greatest philosopher of the twentieth century. I took it on your authority, and that's why I bought another picture of the same man. You may not have *meant* to say it, but because you were careless, you *did* say it." In this case I'm somewhat inclined to side with, and definitely inclined to sympathize with, the philosophically challenged alumnus. I'm inclined to agree that Professor Z did say something that entailed that Agnew was the greatest philosopher of the twentieth century, even though he didn't mean to. Because he was careless, he said something he didn't mean to say.

On the other hand, suppose the audience was composed completely of worshipful philosophy graduate students who know the lecturer and his careless habits well and also know that great philosophers, or at least great logical positivists, don't have slicked-back hair, sharkskin suits, and shyder-like expressions. None of them doubts for a minute that the lecturer intends to refer to C. If after the lecture a student says, pointing at C, "Professor Z said that was a picture of Carnap," I would be somewhat inclined to say he spoke truly. I would surely let the remark pass, if I thought no one in the audience could have misunderstood the lecturer. Still, it seems the student would have spoken more accurately had he said, "It was clear that Professor Z meant to point at that picture [C], and so even though he didn't, and perhaps strictly and literally said that that [pointing to A] was a picture of Carnap, he communicated to all of us, who realized that he was just being careless as usual, that that [C]

was a picture of Carnap." On balance, in this case, it seems that the forensic element in "says" dominates, and would push us toward the demonstrative theory. However, I think for this reason, it isn't quite the concept we need for semantics.

The concept of "what is said" wasn't really designed for the philosophy of language; we need to *explicate* it, as Carnap would say, that is, to replace it with a new somewhat more precise concept, with a somewhat different extension. The official name I'll suggest for the new concept is "locutionary content"—"locutionary" is another term borrowed from Austin, who spoke of the "locutionary act." And I think the concept of locutionary content, in the case of demonstratives, should be tied to directing intentions even in cases, like our second case, where what is said seems, on balance, to go with demonstrations. We need a concept very close to what is said, but one that seals off the forensic issues. (We don't have to use the term "locutionary content" all the time, by the way. The explicator can legitimately usurp the original word, "says" in this case, for serious theoretical work once the explication is finished.) To make this case and develop the concept, we need to look more closely at the structures of beliefs and intentions that govern our uses of demonstratives.

4. REFERENTIAL PLANS AND LOCUTIONARY CONTENT

Let's go back to Case 2, where Professor Z is myopic and the picture of Carnap has been switched for one of Agnew. What does Professor Z intend? It's a mistake to think that we can adequately grasp his intention with a single clause, such as "he intends to refer to the picture of Carnap." We need to see his intentions as forming a structure, a plan, whereby he intends to do one thing by doing another. Professor Z's plan looks something like this. He intends to say, of C, that it is a picture of Carnap. He intends to do that by referring to C, and predicating that property of it.

His plan then includes the subgoal of referring to C. How does he plan to do that? His plan is to refer to C by looking at C and referring to the picture he is looking at. This is where his plan goes awry. The plan depends on the truth of the proposition that the picture he is looking at is C, which it isn't. He does what he intends to do (refer to the picture he is looking at) but not what he intended to accomplish *by* doing that, namely, refer to C. And so, eventually, he doesn't say what he intended to say.

The belief that he was looking at the remembered picture of Carnap was part of Z's motivation for uttering the sentence he did. There were lots of other beliefs that were part of the motivation, in the sense that had he not believed them, he wouldn't have uttered the sentence he did. He had to believe that "Carnap" was Carnap's name, for example. These were not part of what he said, or of what he intended to say. They are not the beliefs he was *expressing* by speaking as he did. The belief he intended

to express was the belief that a certain picture, C, was Carnap. This is a belief he held before he came in the room; he would have continued to hold it even if he had realized that the pictures had been switched. Had he realized that he wouldn't have been under the illusion that he could refer to C by referring to the picture on the wall, and wouldn't have been under the illusion that he could express his belief in the way he tried to do. But he still would have had the belief he was trying to express.

In contrast, the belief that his gesture would pick out C was not part of Z's motivation for uttering what he did. The gesture is not part of his referring, but something intended to help the audience identify that to which he was referring. The careless gesture misled the alumni in a way he was responsible for, and our ordinary forensic concept of what it said is sensitive to this. But I think our semantical concept of locutionary content should not be.

A *prima facie* plausible principle is that if one is sincere, and semantically competent, and makes no verbal slips, what one says will have the same content as the belief one intends to express by saying it. If I sincerely say, "George W. Bush is from Texas," then one can assume that the content of the belief that I was trying to express is that George W. Bush is from Texas.⁴ But, because of the way indexicals and demonstratives work, if for no other reason, we can't really stick with this principle without qualification. Suppose I believe that George W. Bush is from Texas, and I wish to express this belief by saying, "He is from Texas," in a situation in which you and I have just turned on the television late at night, and a very Dubbya-looking fellow has appeared on our screen, making very Dubbya-like expressions and gestures. But it isn't really George W. Bush; it is the George W. Bush mimic that appears every so often on Jay Leno's *Tonight Show*. My plan to express my belief that George W. Bush is from Texas miscarries because I mistakenly thought that in the context we were in I could refer to George W. Bush as "he," exploiting the fact that I was seeing him on the television, but I couldn't.

Now compare this with a case just like it, except that I have actually not looked at the television. I am just going over people in my head and thinking about them. I get to George W. Bush and say, "He is from Texas." I assume you will ask about whom I speak, and I'll say, "George W. Bush." You take me to be, by mistake, referring to the fellow on television, when I meant to refer to George W. Bush. But in this case, it doesn't seem to me that I do refer to that fellow; I refer to George W. Bush, the fellow I was thinking about. My intention was not to exploit

4. Given the international audience for Kaplan's work, perhaps I should point out that George W. Bush, often called "Dubbya," is the president of the United States as I write, and is the son of another president also named George Bush. Since I believe interest in Kaplan's work will endure, perhaps I should observe that the United States is a North American nation that played a leading role in world affairs throughout the twentieth century.

the supposed fact that George W. Bush was on television, for I wasn't looking at the television and hadn't made that mistake. It is really very easy to refer to things. Just think about them, and say "he" or "she" or "it."

In the first case, I intend to refer to the person we both see on television—that is my directing intention—and by doing that refer to George W. Bush. In that case, I don't refer to George W. Bush, don't say what I intended to say, and don't say something that has the content of the belief that I was trying to express. In the second case, my directing intention is to refer to the person I am thinking of, and I do refer to him, I do say what I intended to say, and I do say something that has the content of the belief I was trying to express. (I'm here beginning to extend the concept of directing intention beyond the perceptual cases Kaplan has in mind, which I will continue to do.)

Now here is an intermediate case. I am thinking of George Bush, and that he is from Texas, and decide to express this belief by saying, "He is from Texas," using "he" to refer to the person I am thinking of. Just as I am about to blurt this out, I notice the fellow on television and point to him, as a way of saving you the trouble of asking about whom I am speaking. I don't intend to refer to George W. Bush by referring to the fellow we see on television; I intend to refer to George W. Bush by referring to the person I am thinking of, and then, in addition, help you to identify that person by pointing to him. My gesture has a "uptake-helping" rather than a locutionary purpose. In this case, it seems to me that I say that George W. Bush is from Texas and, in addition, do something intended to be helpful to you in understanding what I say, that is in fact quite unhelpful, although somewhat amusing.

Now perhaps in this case the forensic aspects of our ordinary concept of "what is said" may incline us to think that it is not clear what I said, or even that I said, of the *Tonight Show* comedian, that he is from Texas. Be that as it may, I think for the philosophy of language we want a concept that is closely tied to the actual structure of intentions, and immune to forensic issues; that is, we want locutionary content, that depends unambiguously on the actual structure of intentions, and not at all on the listener's recognition or misrecognition of them, however reasonable these might have been.

The relevant facts about the speaker are his knowledge of the conventions of his language, his beliefs about the things he is talking about, and his identifications and misidentifications of these things with the objects that it possible for him to indexically or demonstratively refer to. These are the factors that enter into the speaker's plan to express, with language, the propositions that he wants to express. If this plan is flawed, because he has made false identifications, then his utterance will not have the locutionary content that he plans on it having, and our *prima facie* plausible principle that sincere competent speakers express the belief they are trying to express is violated. But the problem will be traced to

problems with their beliefs, not mistakes others make in reading their intentions.

Another example. I say, "I think lima beans taste horrible," and intend to emphasize the point by making a horrible expression, but instead make an expression that makes my audience think I am imagining eating a chocolate chip cookie or some other delicious food. They will be very confused. But that's irrelevant to the locutionary content. I expressed my belief but did a very poor job conveying it.

On the other hand, suppose I say, "I think lima beans taste like this," and then make the chocolate chip cookie-eating face, thinking I am making a lima beans-eating face. My intention is to get at the property of tasting horrible *as* the property suggested by the expression I make. But I am wrong about the expression I make. So I don't say what I intend to say. The expression had a locutionary job to do, rather than an uptake-helping one.

I'm now done contrasting locutionary content with our ordinary, semiforensic concept of what is said. Having completed the explication, from now on I will use "says" and its cognates to mean locutionary content.

5. CONNECTIONS AND DIRECTING INTENTIONS

Our ability to think about particular things depends on our having some *connection* to them. Such connections are involved in various kinds of thinking and are exploited when we refer to the things. Examples: I am aware of a pain in my toe; I see the house across the street; I hear an explosion; I hear a person talking; I hear a person talking on the phone; because of earlier perception, I have an idea of a thing and remember it; people are talking about someone, using his name, and I hear the name; someone has written an article about a certain place, and I read the article, perceiving a name for the place; someone refers to something he sees as "that man", and I am aware that he has referred to a man. Each of these ways of thinking about a thing involves having a relation to it, some rather direct, as when we feel our own pains or see objects, some much more tenuous and mediated by the minds and words of others, as when I read about a place I've never visited in a guidebook or hear a snippet of a conversation in which someone is referred to with a name or demonstrative.

If we can think about something, we can refer to it. If a name is available, we can use that. I come upon a conversation in the lounge about someone named Jack. I can wonder who Jack is; I can form hypotheses, which will be true or false depending upon who he is. I can ask, "Who is Jack?" If a name is not available, I can make one up. "Let's call the girl I remember from sixth grade 'Shirley' until we find out her real name." But the easiest way to refer to an object we are thinking about is to use a demonstrative: "that girl," "this pain," "that house," "this guy you call 'Jack'" or simply "he."

In considering what character to assign a simple demonstrative like “that,” we begin by noting that it can be simply a way of referring to the object one is thinking about. If you are thinking about an object, you can refer to that object by saying “that.” You will be referring to it *as* the object you are thinking about. So the appropriate character rule will be: “that” refers to the object that c_A is thinking about at c_T at c_L in c_W . If we could get by with this character rule for “that,” then we would not need to supplement the demonstrative with the directing intention. The referent of “that” in c will be the object about which the speaker is thinking at c_T in c_W .

However, I think that we need to distinguish a number of characters for “that”, corresponding to the types of connections that one can have to objects one thinks about, and choose to exploit in order to refer to those objects. The reason for this is that when one uses a demonstrative (or an indexical or name), part of what one is doing is opening a *second* connection for one’s audience. The first connection is simply that constituted by one’s own remarks. If I say, “He is a fool,” you understand that, if my remark is true, someone is a fool. Who? The person I referred to with “he.” That’s one connection you have to the person about whom I am talking, one way you can think of him. If my remark comes out of the blue, that will be all you can discern about this fellow; you have only one channel of information about him, what I say. But usually you will take my use of “he” to be tied to some *additional* connection that I have to the person, that you can at least potentially use as the basis of a second connection. For example, you will take it that I am referring to the person I am looking at, who you may then look at also. Or you may take it that I am referring to the person who has been the topic of conversation, or the fellow I am talking to on the phone.⁵

Kaplan’s demonstratives take the form “that[δ],” where δ is the directing intention. I suggest that δ be the *type* of directing intention, where the typing is by the type of connection that the speaker is exploiting. A theory of such types could, I think, be based on the concepts of epistemic and pragmatic relations, which I have developed in various places.⁶ But for present purposes I’ll just assume that any theory of demonstratives can incorporate a theory of connection types suitable to the range of discourse treated by the theory.

I am assuming that although it is easy to say something about a thing or object one can think of, and although one can say something (in my explicated sense) with no thought at all about the effects on others, nevertheless language is mostly for communication and provides us the

5. See John Perry, “Understanding Demonstratives,” in Michael Devitt and Richard Hanley, eds., *Blackwell Guide to the Philosophy of Language* (Oxford: Blackwell, 2006).

6. See John Perry, “Rip Van Winkle and Other Characters,” in *The European Review of Philosophy*, Jérôme Dokic, ed., vol. 2, *Cognitive Dynamics*, 13–39; reprinted in John Perry, *The Problem of the Essential Indexical and Other Essays*, enlarged edition (Stanford, Calif.: CSLI Publications, 2000).

resources to say what we have to say in such a way that we can control, to a certain extent, the effects we have on our listeners. Why, then, one might ask, doesn't English come with a variety of "thats," subscripted so that listeners can more easily discern *how* the speaker intends to refer to an object, and so the speaker might more easily control how the listener takes him to be referring? The answer to this question is, I think, that "that" is designed for use without further elaboration when it is pretty obvious how the speaker intends to refer, and easily accommodates further elaboration, in the form of nouns, appositives, and the like, when needed.

One might doubt, on general principles, that our referential plans are so fine-grained as this theory seems to suppose. I suggest some thought and language experiments. Look at something. Refer to it as "that." I'm now looking at the Sprite can on the table, and I just said, out loud, "That is empty." Now I'm not looking at it. I'm remembering taking a full can of Sprite from the refrigerator yesterday. I just said out loud, "That can is now empty." I was referring to the can I referred to as the can connected to me through this memory, not as connected to me through perception. I think it is the same can as the one I can see; I'm almost certain. If later I find a relatively fresh and pristine, but empty, can of Sprite on the desk where I worked yesterday, and inspect the can on the table and find many signs that it has been there for some time—it contains the remains of a couple of cigars, a few discarded bent paper clips, and the carcasses of a couple of dead flies—I will reconsider. It seems that the can I took from the refrigerator yesterday is on the desk; the can on the table is a different one, which I drank from last week and then used as a mini-wastebasket. So I said two different things.

Were they both true? That will depend on what I meant by "empty". Did I mean "contains no Sprite" or "contains nothing at all"? The word "empty" can be used to get at a number of properties, depending on what one intends to say with it, and it can also be used less definitely. If I meant, in my first remark, about the can on the table, that it contains no Sprite, what I said was true. If I meant *completely* empty, in the state required by the city of Palo Alto if I am to toss it in the recycling bin, what I said was false. If I had no intention either way, then I didn't say anything clear enough to have a definite truth value one way or the other.

Suppose I see a man standing across the street. I decide to say something about him. I don't recognize him as anyone I have ever seen or heard of before. He just strikes me as worthy of mention. So I say, "He looks old and tired." Here my referential plan is quite simple. There is a way of referring to the man one sees, no matter who he is. You just look at him and say "He".

A minute later I don't see him any longer. I remember him. I want to refer to the man I remember. Again, it is quite simple. I just say "he." It's a perfectly good way of referring to the man I remember, whoever he was. I may sound stupid, if I say out loud, to no one in particular, "He

looked old and tired." If I am talking to someone, it may not be at all helpful to my communicative goals. The person to whom I am talking may have no idea to whom I am referring. Still, by my lights, I refer to the man I remember. "He" is a device we can use to refer to any male we can think of.

Suppose I mistakenly take the man standing across the street to be George W. Bush. I say, "That man is from Texas." The person I believe to be from Texas is George W. Bush. I intend to refer to George W. Bush and say he is from Texas. I plan to refer to George W. Bush by referring to the man across the street, which I know how to do. I think this will be a way of referring to George W. Bush because of my false belief that George W. Bush is the man standing across the street. In this case I have referred to the man across the street and said he was from Texas. The directing intention is to refer to the man across the street; referring to George W. Bush is the intended result.

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

So, my understanding of directing intentions is this. I think you can use a demonstrative to refer to anything you can think of. It may be something you remember, or see, or hear, a sensation you are having, someone you are talking to on the phone, or whatever. However you think of it, if you can think of it you can form an intention to refer to it, and can do so. If, in the middle of a lecture, I stop, look into the distance, and say, "That man is a fool," intending to refer to some government official, I will have referred to that person, and quite possibly said something true. Reference is easy.

Communication is not so easy. Doing it well requires some skills that most children lack to a certain extent, and some people never acquire. Many children regularly use demonstratives for things that they can see, but which the people they are trying to communicate with cannot identify. My favorite example, because it has happened to me so often, is a child in the backseat of a car I'm driving telling me all about something I can't possibly find without taking my eyes off of the road, turning around, and following their gaze. Many adults regularly use mystery anaphors, where they clearly intend to add to what has been said about some thing referred to earlier, but there is no earthly way of knowing which thing they have in mind. But I digress.

Often one will intend to refer to something one is thinking of, perhaps something one sees or remembers, and intend thereby to refer to an object that has certain further properties, besides being the thing that one sees, remembers, or whatever. One thinks that the thing one sees or remembers or hears is the thing or person that bears a certain name, or falls under a certain description. In this case one intends not only to refer to the thing or person one sees, hears, remembers, or whatever, but also to refer, by doing that to the thing that bears the name or fits the description. Professor Z intends to refer to C by saying "that;" and so forth. The first intention, the one that is easy, is the directing intention. The further referential intentions are fulfilled only if the identifications are correct. What one says, in the explicated sense, depends on one's directing intention. Kaplan is right in "Afterthoughts."

6. RISK AND REFERENCE

Let's return to Professor Z, Case 2. There he is, poor myopic thing. He wants to refer to C, the picture of Carnap. According to me, this is extremely easy for him to do. He has an idea of C; he has seen it many times; he remembers it well. All he has to do is think of it, as something he remembers, and say, "That is a picture of Carnap." He will have definitely said what he wants to say. That seems like the safest way to do it.

Instead, he looks over at the wall and focuses on the picture he sees, however myopically, hanging there. He forms the intention to refer to the picture he sees and thereby refer to C. Why does Professor Z settle on such a risky plan to refer to C, a plan whose success is hostage to his myopic vision?

He settled on the risky plan because his intention to say that C is a picture of Carnap is itself part of a larger plan, to communicate something to his audience. Perhaps he wants his audience to know what Carnap looked like. Or perhaps undergraduates have been defacing the picture, taking it to be merely some nonphilosopher in the university's past, perhaps the donor of the room, or a much-revered groundskeeper from the 1950s; if they realize it is a picture of *Carnap*, they will behave more appropriately toward it. He can't achieve such goals merely by saying that C is of Carnap; he has to say it in a certain way. He has to refer to C in a way that will bring it about that C is playing a certain role in the cognitive life of the audience, so that they can identify C perceptually, and thus find out what Carnap looked like, or find out which picture deserves honorable treatment. It is this aspect of directing intentions, as I am interpreting them, that makes it reasonable to call them "directing"; often they are intentions to refer to an object in such a way that part of the audience's recognition of one's intention will involve directing their attention to an object connected to them (or an object that thus becomes connected to them) in a certain way.

Now it is possible, on my view, for Professor Z to refer to C as the picture he remembers while *pretending* to refer to it as the picture he sees. Then he will say something true, even if he is wrong about what picture he sees; if he is right, his remark will have the desired effect; if he is wrong, and the picture he sees is not C, he will still have said something true, in our nonforensic sense of “says”. This is possible; however, it is not so easy to do. The simplest way to have people take you to be referring to something you see as something you see is to refer to it that way. It would also be rather pointless; Professor Z wouldn’t have spared himself the responsibility for the consequences of his utterance, simply because he managed to say something true in a sense of “says” in which such responsibility is ignored.

Most likely, his directing intention is somewhat impure. He intends to refer to a picture he takes himself to be connected to in two ways, as one he remembers, and one he sees. In that case, it won’t be clear what he said, and there may be no definite answer to the question whether he said something true in a misleading way or something false in a straightforward way.

7. COMPLEX CONNECTIONS

My recommendation for demonstratives within the CCC framework is then as follows. First, we should distinguish among a number of types of connections that one can have to objects: $\delta_1, \delta_2, \dots \delta_n$, guided by the range of cases we want the theory to handle. Then the character of “That[δ_i]” will be a function from a context c to the object that has connection δ_i to c_A at c_T in c_W . Demonstratives that do not get us to an object, even with a directing intention, will need to be given some appropriate treatment.

This way of looking at things might lead in one direction that will be helpful to the *logic* of demonstratives. Our theory of connections might allow for complex connections, which could induce a logic. When we refer to something, we create a connection for our audience; they can think of the object we refer to as the object we refer to. This also provides the speaker with a new connection to the same object just referred to. Once I have referred to an object, both I and my audience can exploit the connection I have created to refer to the same thing again. It is in this way, I think, that we should deal with anaphoric connections—a strategy similar to that adopted by a number of linguists and philosophers of language who do not find Kaplan’s claim that demonstratives and pronouns are ambiguous very inviting.

Consider the sentence “If he waits for a bus, he waits for a bus.” The first “he,” let’s suppose, is a demonstrative with a perceptual directing intention. If the second “he” is anaphoric, and the anaphoric “he” is simply a homonym of the demonstrative “he,” as Kaplan supposes, then we can account for the feel of logical validity, but not as a part of

the logic of demonstratives. If, on the other hand, we consider the second “he” as also a perceptual demonstrative, then the truth of the conditional seems not to be a matter of logic at all but to depend on the fact that the two exploited perceptual connections are of the same fellow.

It seems, however, that anaphors are ways of exploiting connections established by previous directing intentions. We are talking on the phone; you say, “Elwood called again.” I have particular idea to which of several people named Elwood you are referring; perhaps you are referring to an Elwood I don’t even know. I am able to reply, however, in a way that refers to the particular Elwood you referred to. I may ask, “Is Elwood becoming a problem?” I am referring to the same person to whom you referred; I may be trying to discern which Elwood you are referring to—but, equally, I am trying to discern which Elwood I am referring to. Or I may say, “Elwood calls me a lot, too,” making a guess and referring in a nondeferential way to the Elwood I take you to be referring to. Then you may reply, “No, *he* doesn’t call me, I am talking about a different Elwood.” Your “he” refers to my Elwood.

There seems to be room, then, for a theory of complex connections, which a theory of anaphors could use. An anaphoric directing intention points at another directing intention and has the same intended referent as the intention pointed at:

The referent of “that[$\delta_A \rightarrow \delta_X$]” in c = the referent of “that[δ_X] in c .

We would see our problematic sentence as having either the form

If $he[\delta_p]$ waits for a bus, $he[\delta_A \rightarrow \delta_p]$ waits for a bus,

which is valid, or

If $he[\delta_p^1]$ waits for a bus, $he[\delta_p^2]$ waits for a bus,

which is not. In the latter sentence I have used superscripts to distinguish between the two perceptual directing intentions; this means the world of the context will have to have facts in it that provide the possibility of the two directing intentions being intentions to refer to different objects; the sentence will be true at context c if in c_W the object intended by δ_p^1 and that intended by δ_p^2 will be the same.

At this point, however, my ability to both pursue anything that is a plausible interpretation of Kaplan and think about things in my own way sort of runs out of gas. This is because I naturally think in terms of a theory of utterances, so that different utterances of the same demonstrative, each with its own directing intention, anaphoric or not, would occur at different times and would be causally relevant to the production of different tokens of the same word. But Kaplan’s theory is not a theory of utterances, but rather a theory of sentences in context, and the sentences importantly need not be uttered in the world of the contexts in which they are considered. So this is probably a good time for me to stop.

CHAPTER 12

Can Frege Pose Frege's Puzzle?

Stavroula Glezacos

In "Afterthoughts," David Kaplan characterizes Gottlob Frege's "On *Sinn* and *Bedeutung*"¹ theory of proper names as "gossamer;"² nonetheless, Kaplan affirms the commitment at the heart of that theory, writing: "it is indisputable that distinct proper names have distinct cognitive values."³

Frege maintains that two name-containing identity sentences, represented schematically as $a = a$ and $a = b$, can both be made true by the same object's self-identity but nonetheless differ in their epistemic profiles. According to Frege, sentences of the form $a = a$ are knowable a priori and uninformative; those of the form $a = b$ are often informative, and cannot always be known to be true a priori. Frege solves his puzzle by locating a source for this purported epistemic difference between the identity sentences: the distinct *Sinne*, or senses, expressed by the names that the sentences contain.

Thus, Frege portrays himself as describing a puzzle that arises independently of any particular theoretical position regarding names, and then solving that puzzle with the theoretical apparatus of *Sinn* and *Bedeutung*. In this chapter, I will suggest that Frege's presentation is problematic. If an attempt is made to characterize the epistemic status of true identity sentences without appeal to Frege's theoretical commitments, then what initially seemed puzzling, largely dissolves. It turns out that, in order to generate his puzzle, Frege must invoke that which he seeks to establish the necessity of: the theoretical account of *Sinn*.

My case depends on taking seriously a question that Kaplan poses in his paper "Words": What determines a name-containing identity

1. "On *Sinn* and *Bedeutung*" (hereafter "S and B"), in *The Frege Reader*, ed. Michael Beaney (Oxford: Blackwell, 1997), 151–171.

2. Kaplan writes, "It is on the rock of distinct cognitive values for distinct names that Frege erected his gossamer theory." From "Afterthoughts," in *Themes from Kaplan*, ed. Joseph Almog, John Perry, and Howard Wettstein (Oxford: Oxford University Press, 1989), 598.

3. "Afterthoughts," 598.

sentence's form?⁴ If an identity sentence's form is specifiable without appeal to something like the theoretical notion of sense, then Frege will not be able to maintain that sentences of the form $a = a$ are, in general, uninformative or knowable to be true a priori (at least no more so than true sentences of the form $a = b$). If, instead, sameness or difference in *Sinne*, expressed by the names in the sentence, is what determines an identity sentence's form, then the puzzle fails to hold as a general puzzle; it arises only if we assume a role for the very sorts of theoretical entities that Frege presents it as leading us to.

1. THE FORMS OF TRUE IDENTITY SENTENCES

Frege does not pose his puzzle via actual name-containing sentences but, rather, by using two schemas, $a = a$ and $a = b$, to represent the forms of name-containing identity sentences.⁵ Frege places sentences in one or the other category by listing the epistemic features that they do or do not possess—most prominently, a priori knowability. Sentences of the form $a = a$ “hold a priori,” while true sentences of the form $a = b$ “cannot always be established a priori.”⁶ The enumeration of sentences' epistemic features leaves unanswered, however, the basic question of what makes it the case that an identity sentence is of the form $a = a$ as opposed to $a = b$.

An answer that would not be legitimate for Frege to give here is: “It is sameness or difference in sense that determines whether a given identity sentence has the form $a = a$ or $a = b$.” This would be illegitimate because Frege is supposedly presenting a puzzle that can be posed independently of any particular solution that he will provide for it. We must first be puzzled, and then we can seek a resolution for our puzzlement. Thus, it must be possible for Frege to specify what gives a sentence one form rather than another without making use of his theoretical notion of *Sinn*—even if he then goes on to attribute the epistemic differences between sentences of different forms to the difference in *Sinne* expressed by the names that they contain.

The most natural account of what determines an identity sentence's form involves the notion of sameness of name. Without making any theoretical commitments, we can say that a sentence has the form $a = a$ when the same name flanks the identity sign, and $a = b$ when distinct names flank the identity sign.⁷ Thus, posed neutrally, the (purported) puzzle is this: What

4. Kaplan asks, “What determines an utterance to have the form ‘ $a = a$ ’ as opposed to the form ‘ $a = b$ ’?” and notes that “... this issue takes us directly to matters of word individuation.” “Words,” *The Aristotelian Society Supplementary Volume* 64 (1990): 95.

5. “ $a = a$ and $a = b$ are obviously statements of differing cognitive value ...,” “S and B,” 151.

6. “S and B,” 151.

7. In “Afterthoughts,” Kaplan writes, “Note that Frege's initial argument makes use only of the uninterpreted forms ‘ $a = a$ ’ and ‘ $a = b$.’ The distinction between *repetition* of a

is the source of the epistemic difference between true identity sentences that contain a single name twice, and those that contain two names?

Of course, this formulation immediately raises a further question: What makes for sameness of name? Frege's "On *Sinn* and *Bedeutung*" answer seems to be that we have the same name whenever we have the same sign/referent combination. That this is Frege's view is strongly suggested by his well-known footnote B, in which he invokes the notion of "the *actual* proper name 'Aristotle' " [italics mine], and writes that, since "opinions as to the *Sinn* may differ," we can "tolerate" variations in *Sinn* assignment among users of that name, "so long as the *Bedeutung* remains the same."⁸ This passage provides a general characterization of a name as a particular sign for a particular referent, and, in addition, affirms that the Fregean theoretical framework accommodates assignment of different senses to a single name.

Given this account of name individuation and sentence form, we will not find much to puzzle us when we consider true identity sentences. In particular, if an identity sentence's form is determined by sameness or difference in the names that it contains, then there will be no in-principle epistemic divide between sentences of the form $a = a$ and $a = b$.

It is clear that someone may encounter the same name twice and go on to deny the truth of an identity sentence containing only that single name, or that someone may learn only after empirical investigation that an identity sentence containing that single name twice is indeed true.⁹ There is nothing involved in the notion of sameness of name that ensures that language users—even the most competent—will be able to identify and re-identify names when they encounter them, even when they appear in a single identity sentence.

Of course, if a speaker encounters the name "Aristotle" on one occasion, encounters "Aristotle" on a later occasion, and recognizes that the name is the same name on both occasions, then (given the

single name and the use of two distinct names is already sufficient to make the points about cognition even before any examples (or even the notion of *Sinn*) are introduced." (598n. 70). As will become clear, I do not agree with Kaplan's assertion that Frege's "points about cognition" (that a single name repeated preserves its cognitive value, and that distinct coreferential names differ in cognitive value) can be maintained without invoking the notion of *Sinn*.

8. "S and B," 153.

9. Saul Kripke gives just such a case in his "A Puzzle about Belief," in *Meaning and Use*, ed. Avishai Margalit (Dordrecht: Reidel, 1979), 239–283. Kripke's example involves Peter, who comes across the single name "Paderewski" on two separate occasions; as Kaplan describes the case: "Peter first heard Paderewski's name in connection with his musical accomplishments and later heard the name used in connection with Paderewski's political accomplishments and then concluded that there were two different [names] 'Paderewski' " ("Words," 94). Someone so situated may well go on to deny the truth of the sentence "Paderewski is Paderewski"—a sentence that contains the same name twice, and is thus of the form $a = a$. Given such denials, we must conclude that sentences of the form $a = a$ are not always knowable a priori.

characterization of name as a sign/referent combination) she can immediately know that "Aristotle = Aristotle" is true. But notice that the ability to recognize that the name is the same seems to involve the ability to recognize that the *referent* is the same. If this is the case, then an identity sentence of the form $a = a$ is not, in principle, recognizable as true in any way different from a true sentence of the form $a = b$. In both instances, once one knows that the referent is the same, one knows that the sentence is true. If one does not know this, then, regardless of whether the sentence is of the form $a = a$ or $a = b$, one does not know (let alone know *a priori*) that it is true.

Frege himself more or less acknowledges this latter point in his "Aristotle" footnote, albeit by invoking his theoretical notions of *Sinn* and *Bedeutung*. He notes that the name "Aristotle" may be assigned different *Sinne* by different users, which is a very short step away from recognizing that the *same* user could assign different *Sinne* to a single name. Though these *Sinne* determine the same referent, such a person may remain unaware of that fact; if he says "Aristotle was wise," thinking about the teacher of Alexander, and "Aristotle was Greek," thinking about the Stagirite philosopher, he is in no way situated to immediately recognize the truth of "Aristotle = Aristotle"—at least, no more so than someone who has associated these distinct senses with two different names (for example, "Aristotle" and "Ari").

In sum, we ought not accept Frege's initial, supposedly puzzlement-inducing claim that sentences of the form $a = a$ and $a = b$ have distinct epistemic profiles. The truth of a sentence of either form can be denied by a perfectly competent language user; furthermore, for each sentence, recognition of its truth involves recognition that the names—either a twice, or a and b—have the same referent. Even after invocation of Frege's framework, the purported puzzle fares no better. Frege has linked the epistemic status of a name-containing identity sentence to the *Sinne* expressed by the names in the sentence. Since, however, sentence form is determined by name identity, and names are not (at least, not in "On *Sinn* and *Bedeutung*") individuated by assigned sense, it turns out that there is no principled reason to ascribe a special epistemic status to sentences of the $a = a$ form—and thus, there is no puzzle about the epistemic divide between true identity sentences of different forms.

2. "DR. LAUBEN"

Frege revisits the issue of what makes for sameness of name in his later paper "Thought," where he considers a situation in which two friends are discussing the travails of a third man.¹⁰ Frege writes:

10. "Thought," in *The Frege Reader*, 325–345.

Suppose . . . that Herbert Garner knows that Dr. Gustav Lauben was born on 13 September 1857 in NN and this is not true of anyone else; suppose, however, that he does not know where Dr. Lauben now lives nor indeed anything else about him. On the other hand, suppose that Leo Peter does not know that Dr. Lauben was born on 13 September . . . Then as far as the proper name "Dr. Gustav Lauben" is concerned, Herbert Garner and Leo Peter do not speak the same language, although they do in fact designate the same man with this name; for they do not know that they are doing so.¹¹

Here Frege seems to be pulled in two directions. On the one hand, echoing his "Aristotle" footnote, he calls attention to *the* proper name "Dr. Gustav Lauben," which is seemingly specifiable as a single name independently of the assignment of *Sinne*. On the other hand, Frege takes the fact that Peter and Garner do not believe themselves to be speaking of the same man as evidence that they are not speaking the same language, and he suggests that we can better account for their situation by supposing "that Leo Peter uses the proper name 'Dr. Lauben' and Herbert Garner uses the proper name 'Gustav Lauben.'"¹² Frege thus seems to settle on the view that the case of Garner and Peter is interchangeable with one in which name-difference appears on the surface of the participants' utterances. Because Garner and Peters have associated distinct *Sinne* with "Dr. Lauben," they have used two different "Dr. Laubens"; name identity, on this view, is (at least in part) determined by associated sense.

We should now revisit the puzzle and inquire once more whether Frege has, in fact, located a matter of general concern. Recall that the puzzle is posed in terms of sentence form, and that sentence form is determined by the identity or distinctness of the names appearing in the sentence. If a name is in part individuated by the *Sinn* associated with it, then, in order to be puzzled, one would need to be committed to *Sinne* and their role in determining name identity. If we are not so committed, we will find that the puzzle has no hold on us.

3. CONCLUSION

Though Frege's discussion of true identity sentences in the opening sentences of "On *Sinn* and *Bedeutung*" has been thought to present a problem for all who theorize about language, it in fact does not present a puzzle for anyone.

In a framework in which *Sinne* are not invoked as name-individuators, it seems that identification of the referent will be necessary for identification of the name; and thus, all true identity sentences will be on an epistemic par. Even within Frege's framework, if different *Sinne* can be

11. "Thought," in *The Frege Reader*, 333.

12. "Thought," in *The Frege Reader*, 333.

assigned to the constituents of a sentence of the form $a = a$, and the same *Sinn* assigned to the names in a sentence of the form $a = b$, then Frege himself should not hold that there are in-principle epistemic differences between true identity sentences of different forms.

However, if the true Fregean position is that a name-containing identity sentence's form is determined by an individual language user's assignment of *Sinn*(*e*) to the name(s) in the sentence, then what emerges is that his "puzzle" and his solution are in fact of a piece. Only one who holds that *Sinne*—or name user/interpreter assigned cognitive values—play a role in determining a name's identity can maintain that $a = a$ sentences are knowable a priori and uninformative, while $a = b$ sentences lack these features. But in what way is this puzzling? The very same feature determines both the identity of the names in the identity sentence (and hence the sentence's form) and the sentence's epistemic status.

The end result is that non-Fregeans should not be puzzled about true identity sentences, for they have no reason to accept that there is an epistemic divide between identity sentences of different forms. Nor, as we have seen, should Fregeans be puzzled, for it is their framework that both generates the divide and explains it.

ACKNOWLEDGMENT. For helpful discussion of these ideas, I would like to thank Joseph Almog, Roberta Ballarín, Erin Eaker, Avram Hiller, Andrew Hsu, Ralph Kennedy, Amy Kind, Christian Miller, Terry Parsons, Seana Shiffrin, Ori Simchin, and, most especially, David Kaplan.

CHAPTER 13

A New Starting Place for the Semantics of Belief Sentences

Erin L. Eaker

This essay concerns the semantics of *believes that* sentences, that is, sentences like, “Stella believes that Joe Tobacco won the hand.” I argue that semantic theories for *believes that* sentences have been unnecessarily complicated by an unsupported assumption about the primary function of *believes that* sentences, namely, that they relate their subjects to contents of belief in a way that reflects the role that those beliefs play in the subject’s thinking, reasoning, and behavior. This assumption, which I refer to as the *Explanatory Function Assumption* (EFA), leads to a distorted understanding of *believes that* sentences and to the mistaken conclusion that any adequate account of the semantics of these sentences must provide solutions to certain well-known but (I will argue) illusory puzzles. My thesis is that when we reject the Explanatory Function Assumption (and another closely related assumption), the puzzles are no longer puzzles, the semantic account of *believes that* sentences is suddenly much simpler, and the burden for explaining the so-called belief puzzles is put where it should be, that is, on the pragmatic account of how belief descriptions are used.¹

First, the Explanatory Function Assumption:

Believes that descriptions primarily function to relate a rational agent to the content of his attitude, with the content of the attitude given in a way that

1. The conclusions I draw about the semantics of *believes that* sentences—such as that substitution of co-designative expressions in *believes that* contexts is truth-preserving, and that apparent puzzles about belief are to be explained in the pragmatic theory—may call to mind the views of so-called neo-Russellians such as Scott Soames and Nathan Salmon. If the conclusions are similar, the path to them is quite different, since I take as my starting place observations about belief description practices rather than a thesis about the semantics of certain singular terms. Soames and Salmon share certain assumptions with the tradition that make their defense of their respective semantics for belief sentences seem heroic to some of their critics. The present essay provides, I think, a stronger platform for the type of theory they offer. Unfortunately, there is insufficient space in this essay to discuss their views.

captures its role in the agent's reasoning or "cognitive architecture." If there isn't a maximally fine-grained match between the content ascribed and the content believed, the *believes that* description will not be able to play its role in explaining and predicting the behavior of the agent.

The Explanatory Function Assumption is closely related to the tradition of treating forms of psychological description that take a "sentential complement" as the canonical forms for psychological description. I call this the *Canonical Form Assumption* (CFA):

The *believes that* form is the canonical form for belief description. When true, *believes that* sentences *completely* and *accurately* reflect the psychological facts. This is because, by their very form, *believes that* sentences relate agents to belief contents, which contents are expressed by the embedded sentence. When true, *believes that* descriptions reproduce the content of the agent's belief, genuinely presenting us with a piece of the agent's mind, as it were. Other forms of psychological description, such as *de re* forms or descriptions that involve direct quotation of what the agent *said*, are incomplete and do not accurately reflect the psychological facts. Because these latter descriptions focus on the objects about which the agent has beliefs or on the agent's external verbal behavior, they fail to give a complete and accurate description of the psychological facts.

In section 1, I will briefly describe some semantic theories for *believes that* sentences that are clearly motivated by the Explanatory Function Assumption and the Canonical Form Assumption, and explain how this leads to a complication of the semantics. In section 2, I will provide some examples of our everyday psychological description practices. I will explain why these examples motivate a methodological shift in our approach to the semantics for *believes that* sentences, including a rejection of the EFA.

In section 3, I will show how three factors have contributed to the strong hold of the EFA and the CFA. In sections 4 and 5, I will discuss the puzzles that have shaped the semantics of *believes that* sentences and explain the relationship between those puzzles and the EFA and CFA. I will offer new ways of thinking about these puzzles and explain why the issues they raise do not require complication of the semantic theory for *believes that* sentences. I argue that the puzzles themselves motivate a rejection of both the EFA and the CFA.

SECTION 1

I start by looking at some of the semantic theories for *believes that* sentences that manifest the Explanatory Function Assumption and the Canonical Form Assumption. The semantic theories discussed manifest these assumptions in different ways, but all agree that the *truth* of a canonical *believes that* description can depend on a maximally fine-

grained match between the content ascribed to the subject and the content believed by the subject. After all, the belief content ascribed is supposed to be capable of explaining the belief's role in the subject's reasoning or "cognitive architecture." Of course, as is well known, one immediately runs into trouble when translating these assumptions into a semantics for *believes that* sentences. The trouble is how to combine these assumptions about the function of *believes that* sentences with the assumption that in a particular context, utterances of, for example, "Joe Tobacco won the hand," "He won the hand," and "J.T. won the hand" could have the same semantic content. It would seem, in that case, that "Stella believes that Joe Tobacco won the hand" or "Stella believes that J. T. won the hand" would ascribe the same belief content to Stella. But (it is argued) surely Stella could believe that Joe Tobacco won the hand, without believing that J.T. won the hand. Since exchanging co-designative expressions in the "that clause" of a belief description can apparently alter the description's truth value, and its explanatory power as a description of Stella as a rational agent, the semantics of *believes that* sentences must encompass more than the content that "Joe Tobacco won the hand" and "J.T. won the hand" have in common. That common content does not appear to satisfy the EFA or bear out the claims of the CFA. Instead, the *way* that the subject of the report (Stella) *thinks about* Joe Tobacco/J. T. when believing that he won the hand must also be included in some way in the ascription. And this requirement must be encoded in some way into the semantics for *believes that* sentences. What follows is a representative sample of strategies for doing just this.²

Frege of course thought that names themselves have descriptive content—that names express senses or "ways of thinking about" the things the names refer to. Frege would therefore have denied that "Joe Tobacco won the hand" and "J.T. won the hand" have the same semantic content. They express different thoughts, according to Frege, thoughts that contain different ways of thinking about that guy. And so of course Stella could believe that Joe Tobacco won the hand without believing that J.T. won the hand. The two embedded sentences express two different thoughts, and so the two belief descriptions relate Stella to two different beliefs.

In "Quantifying In," David Kaplan provides a semantics for *believes that* sentences according to which the canonical *believes that* descriptions

2. Some "neo-Russellians" such as Soames and Salmon reject this line of reasoning by arguing that since "Stella believes Joe Tobacco won the hand" and "Stella believes that J.T. won the hand" ascribe the same belief content to Stella, they must have the same truth value. They explain the apparent difference in truth value by claiming that the two *believes that* sentences can be used to *pragmatically* convey different information. It is this pragmatically conveyed information that is relevant to explanation and prediction of Stella's behavior. Here I will focus on a representative sample of those who are led by the EFA and CFA to encode the information relevant to explanation and prediction of an agent's behavior into the *semantics* of *believes that* sentences.

are those that function just as Frege says they do, and just as the EFA prescribes. But Kaplan goes further and addresses those belief descriptions that don't appear able to do the work required by the EFA and which don't appear to conform to the CFA. In some *believes that* descriptions—such as those that involve quantification into the belief context—the subject's "way of thinking" might not be preserved in the description. On Kaplan's theory, the underlying logical form of belief descriptions reveals that they *always* involve quantification over the subject's "way of thinking" about the object about which she has a belief. This hidden reference to the subject's way of thinking about the object about which she has a belief is necessary if the belief description is to accurately reflect the belief facts.³

Mark Richard has proposed that *believes that* sentences relate cognitive agents to "Russellian Annotated Matrices" (RAMs). RAMs are composed of objects, properties, and relations as well as the linguistic expressions (or perhaps internal mental entities) that represent them. On Richard's picture, RAMs are what are expressed by sentences, and they are the objects of belief. To believe "Joe Tobacco won the hand" and to believe "JT won the hand" is to be related to different RAMs. Though the RAMs expressed by these two sentences contain the same person and the same property (of winning the hand), they contain different linguistic expressions—"Joe Tobacco" and "JT", respectively. Thus, the two sentences express different RAMs, and to believe one is not necessarily to believe the other. On Richard's view, the predicate "believes" is indexical—in different contexts of utterance, it expresses different relations between subjects and RAMs. To simplify, in some belief description contexts, the RAM ascribed must be very closely correlated with the RAM that the subject herself believes. In other words, in some contexts, the truth of the belief description depends not only on the embedded sentence expressing a RAM that contains certain objects, properties, and relations but also on its being composed of the same words that the subject employed (or would employ) in expressing her belief. In such contexts, in order to serve its purpose for explaining and predicting Stella's behavior, a true *believes that* sentences must relate Stella to precisely the RAM she believes. Richard's inclusion of objects, properties, and relations *as well as* linguistic expressions in the RAMs reflects his assumption that the belief content ascribed must be capable of a maximally fine-grained match with the belief content believed in order to do its work in explaining and predicting the subject's behavior.⁴

3. David Kaplan, "Quantifying In," *Synthese* 19 (1968–1969): 178–214. Kaplan's views on the semantics of belief sentences have evolved substantially over the years. I discuss this evolution in a paper entitled "David Kaplan on *De Re* Belief," *Midwest Studies in Philosophy* 28 (2004): 379–395.

4. For details, see Mark Richard, *Propositional Attitudes* (Cambridge: Cambridge University Press, 1990); Richard, "Attitudes in Context," *Linguistics and Philosophy* 16, no. 2

Another theory that exemplifies the EFA is the Interpreted Logical Form (ILF) theory of Richard Larson and Peter Ludlow. This theory provides semantic contents for the “that clauses” of *believes that* sentences that are maximally fine-grained for capturing the role of a belief content in the cognitive architecture of the subject.⁵ According to this theory, *believes that* sentences relate agents to Interpreted Logical Forms that are expressed by the embedded sentences. Like Richard’s RAMs, these structures contain objects, properties, relations, and the linguistic expressions that designate them. Furthermore, Larson and Ludlow suggest that, though it is an empirical question, ILFs may also contain phonological information. If so, then to say, for example, “Stella believes that [harvard] is highly selective” and to say “Stella believes that [hahvahd] is highly selective” is to relate Stella to two different ILFs.⁶ The inclusion by Larson and Ludlow of all this information in the semantics for *believes that* sentences is motivated by their assumption that *believes that* sentences must maximally reflect maximally fine-grained distinctions between the beliefs ascribed to cognitive agents in order to do the work they are supposed to do, that is, serve in explanations and predictions of the behavior of cognitive agents.

To summarize: All four of these theories adhere to the EFA and the CFA. They all start from the premise that *believes that* descriptions are the *canonical form* for describing the psychological facts because they relate a rational agent to the content of her attitude, with the content of the attitude given in a way that captures its role in the agent’s reasoning or “cognitive architecture”. They think that for the description to do this work, there must be a maximally fine-grained match between the content ascribed to the subject and the content she believes. That is what compels these theories to come up with increasingly complicated semantic accounts—accounts that are capable of taking into consideration all the peculiarities involved in the way a particular subject thinks about the matter in question.

(1993): 123–148; and Richard, “Defective Contexts, Accommodation, and Normalization,” *Canadian Journal of Philosophy* 25, no. 4 (1995): 551–570. Useful criticism of Richard’s view has been offered by Ted Sider, “Three Problems for Richard’s Theory of Belief Ascriptions,” *Canadian Journal of Philosophy* 25, no. 4 (1995): 487–513; and Scott Soames, “Beyond Singular Propositions,” *Canadian Journal of Philosophy* 25, no. 4 (1995): 515–549.

5. Richard Larson and Peter Ludlow, “Interpreted Logical Forms,” *Synthese* 95 (1993): 305–356, reprinted in *Readings in the Philosophy of Language*, ed. Peter Ludlow (Cambridge, Mass.: MIT Press, 1997), 993–1040.

6. Though they admit that it is controversial whether such phonological information is available at the level of “LF,” if this phonological information were present in the phrase-markers, the ILF theory predicts it can give rise to distinct ILFs.

SECTION 2

Now I want to consider some examples of *believes that* sentences that clearly do not function in accord with the EFA.

Example 1

My friend Giovanni has said to me on more than one occasion, “Londra è bella.” Likewise, my friend Philippe has, on occasion, expressed his opinion that, as he would put it, “Londres est jolie.” Suppose I’m in Toronto talking to my friend John, who is trying to decide where to go on vacation in the spring. I say to John: “My friends Giovanni and Philippe, both well-traveled Europeans, believe that London is a pretty city.” Notice that I use the English word “London” where Giovanni and Philippe, the subjects of my psychological description, used “Londra” and “Londres,” respectively. I use the word “London” because that is the word John and I (and Air Canada’s Web site) use to talk about the city London. Neither Giovanni’s nor Philippe’s “ways of thinking about” London are reflected in my description. We can assume that Giovanni and Philippe do not speak English, and this won’t affect the truth of the description. Notice that I also correct for my friend Giovanni’s tendency to overstate his opinions by describing him as believing of London that it is *pretty* rather than *beautiful*. In short, this description does not adhere to the EFA, and yet there is no reason to deny that it is true. Its point is not to explain and predict Giovanni and Philippe’s behavior, but rather to influence John’s. I am trying to convince him to go to London in the spring by appealing to the fact that my well-traveled European friends agree about its attractiveness. In fact, if I wish to convey their agreement, the different ways they refer to the city and think about it *must* drop away.⁷ The violation of the EFA in this case actually serves the purpose of the description.

Example 2

Suppose that despite Dan’s best intentions he has managed to spoil his daughter, so that on Christmas Day, she is sitting by the tree surrounded by wrapping paper and toys and is complaining about what Santa brought her for Christmas. Suppose also that Dan has just gotten off the phone with some French relatives and that his daughter’s cousin Philippe, despite

7. Similarly, as Kaplan pointed out in “Demonstratives,” if I am to give you an accurate account of what Henry believes when he says to me, “You are insufferable,” I must not report to you, “Henry believes that you are insufferable.” Accurate representation of what Henry believes in this case begins with correct reference to the individual about whom Henry has a belief. David Kaplan “Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals” in *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press, 1989), 481–563.

having received only some socks, a book, and *one* new Xbox game, nonetheless thinks it fair to say: “Père Noël est généreux!” Dan seizes this opportunity to teach his child an important lesson about gratitude and says to her: “Well, your cousin Philippe believes that the fellow who brought you a DVD player for your own bedroom, *and* the complete Disney DVD collection, *and* a Barbie cell phone is quite generous!”

Consider: Philippe doesn’t know what Dan’s daughter got for Christmas, so he doesn’t think of Père Noël as *the fellow who brought her the DVD player, and so forth*. So why does Dan use that description rather than “Père Noël”? He use it because he is trying to instill some values into his Anglophone daughter. He is trying to influence *her* behavior, not explain or predict Philippe’s. If he had instead reported, “Your cousin Philippe believes that Père Noël is generous”—reproducing the content of Philippe’s belief in a way that reflects its role in his cognitive architecture—Dan’s attempt at moral education would have had little chance of success.

While such belief descriptions have not gone unnoticed, most of the examples and puzzles that have shaped the semantics of *believes that* sentences ignore the audience-directed aspects of psychological description and focus instead on whether they accurately predict or explain the (usually linguistic) behavior of the subject. As these two examples show, this often misses the point.

Example 3

This last example shows that even when the purpose of the description is the explanation and prediction of the subject’s behavior, the *believes that* description does not have to provide a maximally fine-grained match between the content ascribed and the content believed in order for the *believes that* description to play its role in explaining and predicting the behavior of the agent. The previous examples emphasized that people using *believes that* sentences are often *not* trying to provide psychological explanations of their subjects, and for that reason their belief descriptions do not adhere to the EFA. This example shows that even when an attempt is being made at psychological explanation, the success of the description does not require a maximally fine-grained match between the content ascribed and the content believed.

The setup: Mathilde is gossiping with Roxy about their friend Christie. Mathilde relates a story that Christie told her on the weekend. Mathilde says: “You know how Christie is always complaining about the chauvinist contractor who is redoing her kitchen? Well, Christie took the afternoon off last Friday and got home from work a few hours early. When she walked in the door, she found someone standing at the kitchen table wearing what appeared to be one of her dresses and fussing over a vase of flowers. Well, when this ‘lady’ heard the door open she wheeled around (on Christie’s heels!) and Christie realized that this Betsey Johnson vision

in black was none other than the swaggering nail-gun brandishing hardware store Hercules himself!"

Let's look at this description in detail. Suppose it is common knowledge between Mathilde and Roxy that Christie doesn't know anything about fashion designers, and so she didn't think of the contractor in his outfit as a "Betsey Johnson vision in black." Furthermore, though Christie has often complained to her friends about the overbearing paternalism of her contractor, the description "the swaggering nail-gun brandishing hardware store Hercules himself" is just catty editorializing on Mathilde's part. But despite all of Mathilde's catty editorializing, her description of Christie's realization perfectly captures Christie's astonishment, certainly explains the "cognitive significance" of her discovery, and predicts that she will call her best friend, Mathilde, to tell her about this exciting new development in her home renovation adventure. In short, Mathilde's description of Christie's psychological state perfectly captures Christie's state of mind. In other words, it does everything that a psychological explanation is supposed to do, but it does this without trying to reproduce any alleged fine-grained ways of thinking about objects that compose the alleged content of Christie's belief.

In fact, *that* we count such psychological descriptions as true is not an accident. It is likely that Christie herself would have been incapable of articulating clearly what her "way of thinking about" her contractor was at the time of this realization. Such is often our position as psychological describers—whether we are describing ourselves or others. Furthermore, even when there is a "way of thinking about" an object that a given subject could articulate in words, we third-person psychological describers often have no access to it, and we can adequately engage in "folk psychological description" without it.

So What?

The key question raised by these examples is this: Can one argue from these sorts of examples of belief descriptions to a conclusion about the *semantics* of *believes that* sentences? Those who have been concerned with the semantics of *believes that* sentences have generally acknowledged the existence of such "loose" belief description practices as I have described. For example, Kaplan, in "Demonstratives," cites the following example of what he calls *pseudo de re* reporting: "'John says that the lying SOB who took my car is honest.' It is clear that John does not say 'The lying SOB who took your car is honest.' . . . What justifies this shocking falsification of John's speech? Nothing! But we do it, and often recognize—or don't care—when it is being done."⁸ While acknowledging the existence of these rogue reporting practices, he does not think they are of any

8. Kaplan "Demonstratives," 555n. 71.

significance for the semantics of “says that” sentences. Presumably he would say the same about *pseudo de re* belief description practices. He writes, “I do not see that the existence of the *pseudo de re* form of report poses any issues of sufficient theoretical interest to make it worth pursuing.” On Kaplan’s view, *pseudo de re* reporting is a deviation from the normal function and as such need not be accounted for in the semantic theory.

The “loose” descriptive practices I have described are, however, sometimes *accommodated* in the semantic theory. For example, Richard’s theory represents an “accommodation” within the semantic theory of the “loose” descriptive practices I have described because it makes the standards for the belief description’s fidelity to the RAM “actually believed” by the subject context-sensitive. In contrast, Larson and Ludlow fall in with Kaplan when they state that the “loose” practices that allow infidelity to the ILF “actually believed” by the subject are to be explained by the pragmatic theory. Despite these differences, as explained in section 1, all three theories complicate the semantics of *believes that* sentences in order to reflect the assumption that *true* and *complete believes that* descriptions must have the resources for a maximally fine-grained match between content ascribed and content believed. Why is this? Why, when we go to give the semantics of *believes that* sentences, do we emphasize only one part of the data about how those sentences function in our everyday practices?

SECTION 3

A careful review of the literature on belief sentences reveals that there are three reasons why the Explanatory Function Assumption and the Canonical Form Assumption have had such a strong grip on the semantics of *believes that* sentences.⁹

Semantics and Philosophical Analysis

The first is the traditional relationship between semantics and philosophical analysis. Many have thought that the philosophical analysis of, for example, causation, belief, or modality goes hand in hand with understanding the underlying logical forms of causation sentences, belief

9. Unfortunately, I do not have the space here to provide the detailed account of this history that is necessary to fully substantiate my claims in this section. I provide a more detailed account of the history of belief sentence semantics in my UCLA dissertation, “Reporting Attitudes” (2002). In any case, the claims made in this section about the assumptions common to typical discussions of belief sentence semantics, while not uncontroversial, are sufficiently commonplace to appear here without detailed substantiation. The more controversial claims of this section are instead those that concern how these commonplace assumptions have led to the distortion of our understanding of *believes that* sentences.

sentences, or modal sentences, respectively.¹⁰ It is not always clear in any particular case whether the metaphysical cart is pulling the semantic horse or vice versa. Regardless, what *is* clear is that *if* you think that doing semantics of belief sentences is tantamount to doing philosophical analysis of belief, you will only be interested in those belief descriptions that reflect the concerns of a philosophical theory of belief. And philosophical theories of belief are concerned with things like what role beliefs play in the subject's rational thought and behavior. Thus, the focus will be on just those belief descriptions that seem to have as their purpose the explanation and prediction of the subject's behavior. Other belief sentences—which seem to have other purposes and which therefore do not produce a close match between content ascribed and content believed—will be ignored. The EFA falls out of this bias toward those *believes that* sentences that seem to actually be geared toward explaining the rational thought and behavior of their subjects.

The other result of this traditional connection between the metaphysics of belief and the semantics of belief sentences is that philosophical theories of belief get reflected in the semantics assigned to the sentences. For example, if you think belief is a relation to a third realm Thought (as did Frege), these Thoughts will be components of the meaning you assign to *believes that* sentences.¹¹ When you impose a philosophical theory of belief onto a semantic theory of belief sentences, you treat the English verb “believes” as if it were a technical term in a philosophical theory of belief. This reinforces the tendency to overlook those belief description practices that don't focus on explanation and prediction of the behavior and reasoning patterns of rational agents. It isn't just that such descriptive practices are used to *do* something else. The question of what a belief description is being *used to do* is not really relevant. When the expression “believes” occurs in a philosophical theory of belief, it represents the metaphysics of the belief relation. That is, it represents *what belief is*, not what belief describers do with belief descriptions.¹²

The *believes that* form has been traditionally thought to provide the most accurate reflection of the belief relation, where belief is understood as a relation between a believer and a belief content, which content is expressed by the embedded sentence of a *believes that* sentence. Other

10. Consider, for example, how the logical forms Quine assigns to sentences containing the expression “necessarily” reflect his metaphysical position on the meaninglessness of *de re* necessity. Likewise, the logical forms Quine assigns to belief sentences reflect his rejection of intentional entities. See, for example, Quine, “Three Grades of Modal Involvement,” in *Ways of Paradox* (New York: Random House, 1966); Quine, “Quantifiers and Propositional Attitudes,” *Journal of Philosophy* 53 (1956): 177–187.

11. G. Frege “Über Sinn und Bedeutung,” *Zeitschrift für Philosophie und Philosophische Kritik* 100 (1892); Frege “The Thought: A Logical Inquiry,” *Mind* 65 (1956): 289–311.

12. David Kaplan's work provides an excellent example of the complex interplay between the metaphysics of belief and the semantics of belief sentences. For example, in “Quantifying In,” the logical form Kaplan assigns to belief sentences acknowledges the

forms of belief description—such as those that relate believers to “external” objects, properties, and relations, or that simply quote what the believer said—do not accurately reflect the metaphysics of belief. When providing a semantic theory of belief sentences is seen as tantamount to providing a metaphysical theory of belief, accepting these other forms of description as legitimate will be seen as tantamount to adopting a questionable metaphysics of belief. And few presently adhere to a metaphysical theory of belief that identifies belief with sincere assertion of sentences; or that considers the belief relation to be exhaustively described by a relation between a believer and “external” objects, properties, and relations; or that claims that there is nothing more to belief than our practices of ascribing them to ourselves and others. Thus, the long association between philosophical analysis and semantics has led to both the EFA and the CFA.

Semantics of Belief Sentences and the “Folk Psychological Theory”

Admittedly, it has become less common in recent years for philosophers of language to think of semantics as resulting in a comprehensive metaphysical map of a given domain of inquiry. Perhaps one reason for this is that the alternative semantic theories that have been offered within any given domain of inquiry conflict so sharply with one another, that the idea that one is discovering rather than imposing a metaphysical view when studying the logical relations among sentences in a given domain seems highly dubious. While philosophers still hope to discover insights about time from the study of tense, or insights about necessity from the study of modal locutions, or even insights about belief from the study of belief descriptions, there is now more emphasis on looking at *all* of the linguistic phenomena relevant to a given domain and in focusing on language *as it is used*.¹³ One might expect, therefore, that semantic theories for belief sentences would become more inclusive with respect to belief description practices. One might expect to see belief description practices aimed at explanation and prediction of a *subject's* reasoning and

existence of *de re* forms of belief description but reveals the essentially Fregean metaphysics of belief that they obscure. Kaplan's view that there is a distinction within the metaphysics of thought between intentional relations that are merely denotational and those that involve a causal connection between an agent, the object about which he thinks, and the representation he uses to think about it is also encoded into these logical forms. Likewise, the logical form assigned to belief sentences in Kaplan's later work “Demonstratives” reveals his shift toward thinking that singular propositions can be the contents of our thoughts.

13. I do not mean to suggest that there is a sharp difference in ultimate motivation or even methodology between classic and more recent philosophy of language. Classically, philosophers of language always looked to language for metaphysical insight. Contemporary philosophers of language simply have at their disposal a more comprehensive description of the language, supported by the research of empirical linguistics.

behavior, and belief description practices aimed at influencing an *audience's* reasoning and behavior treated on a par, with the semantic theory neutral between such practices. But, interestingly, there remains a bias toward those belief description practices that are aimed at explanation and prediction of a subject's reasoning and behavior.

This bias—which provides a second explanation for the prevalence of the EFA—results when one approaches belief sentences not with a prior metaphysical theory of belief in hand but with the aim of discovering the “folk psychological theory” used by ordinary people in their interactions. By hypothesis, the “folk psychologist” is trying to *explain* and *predict* the behavior of his fellows. If you are interested in discovering the psychological theory the folk are using, you are going to be interested in those belief description practices that are geared toward explaining and predicting the behavior of their subjects. As a result, the belief descriptions that are deemed *canonical* will be those that are equipped to do just this. And here again emerge the EFA and the CFA—*believes that* descriptions relate a rational agent to the content of his attitude, with the content of the attitude given in a way that captures its role in the agent's reasoning or “cognitive architecture.” Other forms of report will be set aside because they are not geared toward explaining and predicting a subject's behavior.

Theory of Meaning and Theory of Thought

The third factor that has led to the strong hold of the EFA and especially of the CFA is the relationship that has traditionally been thought to hold between semantic content and thought content. Frege got semantic theory under way with an assumption that has persisted in different guises—the idea that a theory of meaning must simultaneously be a theory of thought. This assumption is still widespread, with philosophers endeavoring to provide a theory of linguistic content that can double as a theory of thought content. If one's semantic theory generates thought contents for sentences of the language, when those sentences are embedded in *believes that* sentences, it is natural to analyze the *believes that* sentence as expressing a relation between the subject and the thought content expressed by the embedded sentence. (The trick, as we saw earlier, is sufficiently enriching those thought contents so that they fully represent the role the thought plays in the agent's cognitive architecture.)

To summarize section 3: The EFA and the CFA fit with some deeply pervasive views about the relationship between semantics and philosophical analysis, especially the philosophical analysis of thought. In what follows, I will continue to challenge both the EFA and the CFA directly and show why these two assumptions should not influence the semantics of *believes that* sentences. I will also explain why none of the preceding three considerations actually lends support for these two assumptions in the face of the challenges I raise.

ENTER THE PUZZLES: INTRODUCTION TO SECTIONS 4 AND 5

As I mentioned earlier, it is generally taken as a given that any adequate account of the semantics of *believes that* sentences must be able to explain away certain long-standing puzzles. The increasing complexity of the semantics for *believes that* sentences illustrated by the four theories I presented in section 2 is due to the creation of ever more sophisticated “puzzles” about belief. The basic criterion for success for any semantic theory for *believes that* sentences has been whether it can handle these puzzles. All these puzzles assume the EFA. They furthermore assume that *believes that* descriptions that reflect the EFA are *the canonical descriptions* for psychological description—they are the descriptions that give accurate and complete descriptions of cognitive agents. While the EFA plays a major role in creating the puzzles, the CFA determines what counts as a solution to the puzzles. Together they lead to the conclusion that if we do not sufficiently complicate the semantics of *believes that* sentences, we will not be able to provide accurate and complete descriptions of cognitive agents, and intolerable falsehoods and contradictions will result.

SECTION 4

A much-discussed puzzle comes from Kripke’s paper “A Puzzle about Belief.”¹⁴ It is a good starting place because the “puzzle” is generated in natural belief description scenarios as a result of principles that are meant to be (and this is crucial) *descriptive of* everyday psychological description practices. They do not result from the employment of a substitutivity principle, by substitution of either semantically equivalent terms or semantically equivalent *that* clauses.

The Puzzle

In Kripke’s paper the following scenario leads to an apparent paradox: Pierre, a Frenchman who speaks only French, has heard of a city that he calls, like his compatriots, “Londres.” He has all the hallmarks of a competent user of this expression as a name of the city London. On the basis of what he has heard, he sincerely and reflectively assents to “Londres est jolie.” A fellow French speaker might describe him as

14. Saul A. Kripke, “A Puzzle about Belief,” in *Meaning and Use: Papers Presented at the Second Jerusalem Philosophical Encounter*, ed. A. Margolit (Dordrecht: Reidel, 1979), 239–283; reprinted in *Propositions and Attitudes*, ed. N. Salmon and S. Soames (New York: Oxford University Press, 1988), 102–148. Page references are to the latter volume.

follows, employing what Kripke calls “disquotation”:

Pierre croit que Londres est jolie.

The standard translation of this sentence into English is:

4.1 Pierre believes that London is pretty.

Here I want to emphasize that, disquotation and translation “principles” aside, any English speaker who understood Pierre’s French assertion would describe Pierre as in 4.1 to a fellow English speaker.

Later on, Pierre finds himself in an ugly part of London and does not know that he is in the city he heard about while in France. Since the people around him do not speak French, he eventually learns English by “direct method” and learns to call the city he lives in “London.” As a now competent English speaker, he sincerely and reflectively assents to “London is not pretty.” Again, normally we would not hesitate to report Pierre:

4.2 Pierre believes that London is not pretty.

Kripke considers whether there is any good reason to refrain from “disquoting” either of Pierre’s French or English assertions. Note that it is not the case that when Pierre travels to England he “changes his mind” about London. He is just as willing as ever to assert “Londres est jolie.” To Kripke, it seems that we must accept both “Pierre believes that London is pretty,” and “Pierre believes that London is not pretty” as perfectly warranted descriptions of Pierre in their respective scenarios. Yet, what happens when we take them together? Kripke stipulates in the beginning of the article that these descriptions of Pierre are to be read *de dicto*, that is, “with the material after ‘believes that’ giving the content of a belief of Pierre.” On this reading, the descriptions ascribe *conflicting* belief contents to Pierre, though he is clearly not irrational. This is the puzzle. Kripke has the following things to say about it:

It is no solution in itself to observe that some *other* terminology, which evades the question of whether Pierre believes that London is pretty, may be sufficient to state all the relevant facts. I am fully aware that complete and straightforward descriptions of the situation are possible and that in this sense there is no paradox. Pierre is disposed to sincere assent to ‘*Londres est jolie*’ but not to ‘London is pretty.’ He uses French normally, English normally. Both with ‘*Londres*’ and ‘London’ he associates properties sufficient to determine that famous city, but he does not realize that they determine a single city. . . . We may even give a rough statement of his beliefs. He believes that the city he calls ‘*Londres*’ is pretty, that the city he calls ‘London’ is not. No doubt other straightforward descriptions are possible. No doubt some of these are, in a certain sense, *complete* descriptions of the situation. . . . But none of this answers the original question. Does Pierre, or does he not, believe that London is pretty?¹⁵

15. Kripke, “A Puzzle about Belief,” 122–123.

Kripke's Demand

It is clear that Kripke doesn't think that the puzzle can be solved simply by giving a description of Pierre's cognitive situation in some other terminology besides the *believes that* form. But why not? Why are the descriptions of the situation in other terminology (which Kripke admits are in some sense "complete descriptions of the situation") not allowed to defuse the puzzles? Who, upon hearing the story, is really puzzled about Pierre's cognitive situation? Kripke himself says that in some sense there is no puzzle, since there are perfectly good descriptions of Pierre's situation in other terminology. And yet, he does not explain why none of these descriptions provides an adequate answer to the original question: "Does Pierre or does he not believe that London is pretty?" In other words, Kripke does not explain his tacit assumption that the burden for representing Pierre's cognitive situation must be borne entirely by the *believes that* sentence form, when we have numerous other means at our disposal for saying what Pierre believes.

Consider Kripke's further claim: "The situation of the puzzle seems to lead to a breakdown of our present practices of attributing belief and even of indirect quotation. No obvious paradox arises if we describe the same situation in terms of Pierre's sincere assent to various sentences, together with the conditions under which he has learnt the name in question."¹⁶ Our present practices of attributing belief can be viewed as broken in this situation only because we have limited ourselves to the *believes that* form of description and to a particular *de dicto* reading of this form. If we were to allow ourselves to describe Pierre's situation as Kripke does in setting up the puzzle, we would not be faced with paradox. But Kripke stipulates that we cannot appeal to such forms of description, and he insists that descriptions 4.1 and 4.2 of Pierre's cognitive situation represent the breakdown of our practices of belief attribution using the *believes that* form, since, *read de dicto*, these descriptions portray Pierre as irrational. Kripke thus stipulates and insists on, in essence, the EFA and the CFA. Together they generate the puzzle and lead to the conclusion that the semantics of *believes that* sentences must involve more complexity than meets the eye. I'll show what is wrong with each claim in turn.

Recall that Kripke stipulates that the *believes that* form is to be read *de dicto*, that is, with the material after "believes that" giving the *content* of Pierre's belief. By stipulation, then, "Pierre believes that London is pretty" and "Pierre believes that London is not pretty" would seem to ascribe conflicting belief contents to Pierre—the sort of belief contents that only an irrational person could simultaneously believe. But Pierre is not irrational. So something has gone wrong—a "breakdown" of our

16. Kripke, "A Puzzle about Belief," 131.

practices of belief ascription. However, this dire conclusion is a direct result of the *de dicto* reading of the *believes that* form, which is in turn based on the EFA and the CFA. Suppose we were to suspend the theory-embedded and unargued-for *de dicto* reading of “Pierre believes that London is pretty” and “Pierre believes that London is not pretty” and ask naively what each sentence says. Naively, “Pierre” refers to Pierre, “London” refers to London, “is pretty” refers to the property of being pretty, and “is not pretty” refers to the property of being not pretty. Thus, the first sentence says of Pierre that he believes of London that it has the property of being pretty, while the second sentence says of Pierre that he believes of London that it has the property of being not pretty. This is, as far as it goes, an accurate reflection of Pierre’s cognitive situation—he believes, of a single city, both that it is pretty and that it is not pretty.

Of course, there could be numerous explanations for this—perhaps Pierre is ambivalent; perhaps he thinks some parts of London are pretty and others are not and could go on to explain the subtleties of his opinion; perhaps he is in the situation Kripke describes. Or, perhaps he is irrational. All these situations might be described initially as “Pierre believes that London is pretty and that London is not pretty.” And this description would be correct *as far as it goes*. We would want more information, perhaps, if that is all we were told, but any of the preceding explanations of Pierre’s cognitive situation would be consistent with what we had been told so far. Absent the assumption that the *believes that* form is the canonical form for psychological description because it relates a subject to a belief content in a way that reflects the belief content’s role in the subject’s cognitive architecture, there is no reason to think that the preceding description unequivocally portrays Pierre as irrational when he is not. And so there is no reason to think that our practices of belief attribution using the *believes that* form have broken down. This would be the case only if the EFA were true, that is, if the truth of a *believes that* description really did require that it provide a complete description of Pierre’s belief and that it do so by reproducing the “content” of Pierre’s belief in such a way as to reflect its role in Pierre’s cognitive architecture. That both the descriptions generated in Kripke’s scenario are warranted by our normal description practices and in fact give a true reflection of Pierre’s cognitive situation—as far as they go—demonstrates that the EFA is false. This possibility is considered neither by Kripke nor in any of the discussions of Kripke’s puzzle with which I am familiar. Discussion of the puzzle often center on the disquotation principles that are employed in deriving the paradox, while the stipulation that the resulting descriptions are to be read *de dicto* goes uncontested. The assumption that there is a proper *de dicto* reading of *believes that* sentences is not challenged.

What has also gone unnoticed in discussions of Kripke’s puzzle is that it undermines the tacit, yet pervasive, CFA. In contrast with the assump-

tion that the *believes that* form is the canonical form for psychological description, here we see that the best way of giving a complete description of what Pierre believes is by reference to his initial physical location and his change of location, the linguistic expressions he uses to express his beliefs when occupying those locations, and some information about his linguistic competence with French and English. As Kripke admits, such a description does not lead to paradox. And yet, he refuses to admit that such a description answers the all-important question: “Does Pierre or does he not believe that London is pretty?” He insists, that is, that we be able to give a complete reflection of Pierre’s cognitive situation using the *believes that* form, with the *believes that* form understood, as Kripke stipulates, as relating a subject to the *content* of his belief, this content being expressed by the embedded sentence.

So long as the EFA and the CFA are in play, it will be objected that any appeal to other forms of psychological description in an attempt to provide a complete and accurate description of Pierre’s cognitive situation merely sidesteps the puzzle by not answering Kripke’s question: “Does Pierre or does he not believe that London is pretty?” In fact, *typical* “solutions” to Kripke’s puzzle address this challenge head-on. The typical “solution” to Kripke’s puzzle will begin as follows: There is some difference in information associated with the two names—there is some difference in how Pierre thinks of the city when he thinks of it with one name and when he thinks of it with the other. This way of thinking may be linguistic, metalinguistic, perceptual, conceptual, and so on. The differences in the ways of thinking are such that Pierre is not rationally culpable for failing to recognize that they are ways of thinking about the same city. If α and β represent these distinct ways of thinking, Pierre cannot be expected to infer from α is F that β is F . On some solutions, the ingenuity comes in in explaining where Pierre’s α and β got lost—that is, in explaining why one or the other of 4.1 or 4.2 is false, despite seeming warranted and in keeping with ordinary belief description practices. On other solutions, 4.1 and 4.2 are provisionally accepted, and the ingenuity comes in in trying to make the logical forms of these two *believes that* sentences reflect the “real” cognitive facts—that is, that Pierre believes that α is pretty and he believes that β is not pretty.

What Is Wrong with Kripke’s Demand

Both of these solution strategies assume that we must accede to Kripke’s demand that the puzzle will only be solved by a satisfactory answer to the question “Does Pierre or does he not believe that London is pretty?” and that the answer must be given in the *believes that* form. But why should we accede to this demand? No argument is given for the assumption that the burden for complete psychological description must be borne by the *believes that* form. We have been given no reason to think that the *believes that* form in particular is best suited for explanation and prediction of

reasoning and behavior; no reason to think that all the tools for explanation and prediction of reasoning and behavior should be found in either the surface or the hidden structure of the *believes that* sentence. This CFA is not, to my knowledge, ever argued for explicitly. Perhaps it is supported nonetheless by the sorts of considerations I raised in section 3. I'll address that possibility now.

In section 3, I suggested that the EFA and the CFA are both supported by the fact that semantic theories have commonly been thought of as theories of thought. Since Frege's famous Hesperus/Phosphorus puzzle, it has been common to assume that the semantic content assigned to a sentence must be capable of explaining that sentence's cognitive significance. On such a semantic theory, semantic contents should be fine-grained thought contents. And when sentences are embedded in *believes that* descriptions, they bring their fine-grained contents with them. So the *believes that* form is ideal for representing fine-grained belief contents. This tidy picture runs aground, of course, on a direct reference theory for singular terms: If the sentences "Joe Tobacco won the hand" and "J.T. won the hand" express the same singular propositions, how can we distinguish between the beliefs ascribed by "Stella believes that Joe Tobacco won the hand" and "Stella believes that J.T. won the hand"? Suddenly our canonical *believes that* descriptions don't seem semantically equipped to make the fine-grained distinctions between belief contents required of them on the EFA.

Curiously enough, this has been used as a chief argument against direct reference theories of singular terms, rather than as a reason to question the EFA or the CFA. While direct reference theory for proper names seems to provide a good reason on its own to question the EFA and the CFA, two other reasons have been presented in this essay. Using example 3 in section 1, I argued that the explanatory function of belief descriptions can be performed without faithful reproduction of fine-grained belief contents. Mathilde's description of Christie's "shocking realization" about her contractor perfectly captured its role in Christie's cognitive architecture despite not attempting to reproduce any alleged fine-grained belief content grasped internally by Christie. Furthermore, in the present section I have argued that in order to get a complete description of what Pierre believes we are forced to abandon the *believes that* form of description in favor of other forms of description, forms that relate Pierre not to belief contents but to places, times, languages, expressions in those languages, and so forth. The claim that *believes that* descriptions provide accurate and complete descriptions of psychological agents while other forms of report do not is contradicted by Kripke's puzzle, where the way to describe Pierre's psychological situation without leading to puzzlement is to abandon the *believes that* form in favor of other modes of description. If such descriptions paint the least puzzling picture of Pierre's situation, then we need to be told why such descriptions fail to be complete and accurate.

In section 3, I also suggested that the CFA might naturally arise from the close historical connection between philosophical analysis of belief and the semantics for *believes that* sentences. *Believes that* sentences are considered to be the canonical form for belief description because they are thought to accurately reflect the alleged structure of the belief relation—a relation between a believer and a belief content or proposition. On this view, other forms of belief description—*de re* or quotational—do not accurately reflect the (alleged) metaphysical facts about belief. Defenders of Kripke's demand on these grounds must argue for two claims: First, they must argue that the semantics of belief sentences should reflect the metaphysics of the belief relation. Second, they must argue that descriptions that relate subjects to objects, locations, properties, relations, and sentences they assert and which provide information about the subjects' competence with a certain language nonetheless fail to provide complete and accurate descriptions of those subjects' psychological states, including *what they believe*.

In the previous paragraph I argued that the second claim (and its flip side, the CFA) is simply unsupported and that we have good reason to think it is false. The first claim is also not trivial: Though conflation of metaphysical analysis and semantics is common for belief, it is not common for other verbs. For example, though there may be a "way of walking" that Stella has when she walks into the room, just as there may be a "way of thinking about" the room when she thinks of it as needing a paint job, this way of walking does not show up in the truth conditions or the logical form of "Stella walked into the room." There is an argument for why "ways of thinking" must show up in the semantics of *believes that* sentences, even if "ways of walking" do not show up in the semantics of *walks into* sentences. I address this argument in section 5.

SECTION 5

At this point it may be objected that I have been talking as if it is only tacit assumptions about the relationship between semantics and philosophical analysis or theory of thought that could lead someone to accept the EFA and the CFA and to think that the semantic theory for *believes that* sentences is very complicated. But, it may be objected, do not puzzles that arise from the failure of substitutivity of co-designative expressions in *believes that* sentences by themselves motivate the complication of the semantics? Indeed, failure of substitutivity puzzles are precisely the justification appealed to for including "ways of thinking" in the truth conditions of sentences that contain the verb "believe" even though "ways of walking" don't show up in the truth conditions of sentences that contain the verb "walk."

I propose that we look more carefully at how failure of substitutivity puzzles are developed. Failure of substitutivity puzzles often begin with a seemingly naive prompting such as, “Isn’t it possible for John to *believe that* Hesperus is luminous without *believing that* Phosphorus is luminous.” We are supposed to reply, “Yes, of course it is possible for John to *believe that* Hesperus is luminous without *believing that* Phosphorus is luminous.” This could happen if, for example, John calls the planet he sees in the early morning sky “Phosphorus” and calls the planet he sees in the early evening sky “Hesperus” and doesn’t realize that they are the same planet, Venus.

If this is possible, then it is argued that the semantic theory has to respect certain intuitions concerning the following sentences:

- 5.1 John believes that Hesperus is luminous but he does not believe that Phosphorus is luminous.
- 5.2 John believes that Hesperus is luminous but he does not believe that Hesperus is luminous.
- 5.3 John believes that Phosphorus is luminous but he does not believe that Hesperus is luminous.

It is argued that 5.1 must be assigned a meaning so that it comes out noncontradictory and true. It must, furthermore, be nonequivalent to 5.2, which is supposed to be contradictory and therefore false. It should also be nonequivalent to 5.3, which, it is argued, is not contradictory but also is not a true description of John’s cognitive state. The semantic theory will have to have some way of semantically distinguishing these descriptions in order to preserve our intuitions about their respective truth values. This will depend on finding some appropriate way of including John’s distinct representations of Venus in the logical forms of the respective psychological descriptions.

But this will be difficult, since there are good reasons to suppose that in normal contexts, the names “Hesperus” and “Phosphorus” have the same semantic content, namely, the planet Venus. In the context of sentences that are not used for psychological description, it is common to suppose that names simply refer to their referents, and that the referents themselves exhaust the semantic content of the names, that is, *what is said* by means of them. With this presupposition about the functioning and semantic contribution of the names “Hesperus” and “Phosphorus” in 5.1, the description says, to state it concisely, that John both does and does not believe that a particular object has a certain property. This is exactly what 5.2 and 5.3 say too. Is this a contradiction or a true description of John’s cognitive state? On the semantic theory under consideration, all three of these sentences are contradictory, so we should not say that any of them provides a *true* description of John’s cognitive state. But the interesting thing about these sentences is not their semantics, which, we are assuming, reveals them to be contradictory. The interesting thing is that we could find ourselves asserting one of these

sentences even if it is contradictory. We do sometimes assert things that are contradictory, or necessarily false, like “Hesperus is not Phosphorus” or “I am not myself today.” Sometimes we answer in response to a question: “Well, it is and it isn’t,” or “She does and she doesn’t.” Sometimes saying contradictory things is a good way of conveying something else. So the interesting thing about these psychological descriptions is explaining why we would ever assert any of 5.1 through 5.3 even if what they say is literally contradictory.

I’ll say something about that in the process of answering an objection that goes as follows:

Hold on—you’ve missed the whole point! If your semantics makes these three sentences come out equivalent, then your semantics is just wrong. When philosophers complicate the semantics of *believes that* sentences by appealing to John’s distinct ways of thinking about Venus in the logical form of 5.1, this is not an ad hoc strategy or merely the product of a metaphysical theory of thought. Instead, it is merely an extension of your much vaunted everyday psychological description practice. That practice says that 5.1 is a true description of John’s beliefs because it preserves his actual ways of referring to Venus on the two occasions and therefore provides a complete and accurate representation of what John believes. This description provides the contents of John’s beliefs in a way that reflects the role they play in John’s cognitive architecture. Because 5.2 and 5.3 fail to do this, they are false. If this is the case, it would seem to follow that the way in which the object of John’s conflicting ascription is picked out *does* in fact matter to the *truth* of the descriptions, and thus that something like the Explanatory Function Assumption must be operative in our psychological description practices.

This objection encapsulates the argument for complicating the semantics of *believes that* sentences by including “ways of thinking” (senses, “ways of believing,” RAMs, ILFs, etc.) in their logical forms. Because I have emphasized the importance of paying careful attention to our everyday psychological description practices, I must give proper weight to the intuitions behind this objection. The key intuition that supports the objection is that 5.1 is true *because it reproduces John’s modes of reference to the planet Venus on the two occasions and therefore provides a complete and accurate representation of what John believes*. I would like to see how this intuition holds up outside of failure of substitutivity puzzle contexts and in the context of actual scenarios in which 5.1, 5.2, and 5.3 might be employed.

Suppose that a speaker produced any of these three descriptions of John in some context *and* suppose that the hearer in this context knows that “Hesperus” and “Phosphorus” refer to the same planet. This knowledge will play a role in the hearer’s interpretation of these descriptions of what John believes. The hearer thus starts with the assumption, based on his understanding of the meanings of the expressions that make up the sentence he has just heard, that what has been said is that John believes

and he does not believe that a certain object is luminous. But this appears to be a contradiction. So the hearer must figure out why the speaker would utter this sentence.

The hearer will also be struck by the repeated reference to the same object in the same sentence using two names (or, in the case of 5.2, the same name). This is unusual. When you want to predicate two things of Venus, or you want to say that John believes Venus has each of two properties, or that John believes Venus has one property but does not believe it has another property, you would normally refer to Venus only once in doing so. You might say, for example, "John believes that Venus is a planet and is closer to the Sun than the Earth is." Or, if you referred to Venus twice, you would use an anaphoric pronoun for the second reference. For example, you would say, "John believes that Venus is a planet and that *it* is closer to the Sun than the Earth is." The repeated reference to the same object in 5.1, 5.2, and 5.3 by the same or distinct *names* opens up another interpretive question for the hearer. So, contrary to the typical assumption that 5.1 is an exemplary case of a canonical *believes that* description functioning in accordance with the EFA, 5.1 actually presents two interpretive problems for the hearer. How the hearer resolves these interpretive problems depends on what he knows or believes—about John, about the speaker, about the speaker's knowledge and beliefs about John and his interlocutor (the mutual beliefs and knowledge that go into what is called "common ground").

Here is one way the hearer might resolve that interpretive problem: Upon hearing 5.1, the hearer may *conclude* that the *speaker* knows that John has, in ignorance, expressed conflicting attitudes toward a single object, and that the speaker uses the two names "Hesperus" and "Phosphorus" in his description in order to communicate something concerning how this came about. In this case, the hearer may draw the conclusion that "Hesperus" and "Phosphorus" are being used quasi-quotationally, that is, they are being used not only to refer to the planet Venus but are themselves being mentioned in order to convey something about John's perspective on the world. What is important to note is that this interpretation results *only* because the belief description stands out as unusual given that the hearer knows that "Hesperus" and "Phosphorus" refer to the same individual: It is unusual to refer twice to the same object by different names in the same sentence, and it is unusual to assert something that is contradictory. This causes the hearer to wonder what the speaker intends to convey by means of the repeated reference using different names, and may lead him to interpret the description as attempting to convey how John's perspective on the world has led him to his conflicting attitudes about the planet Venus. The speaker, for his part, is exploiting the norms governing the assertion of contradictions and

the norms governing repeated reference to a single object within a single “turn” in a discourse, to elicit this interpretation.

The hearer would have to embark on a similar interpretive exercise in order to understand an assertion of either 5.2 or 5.3. The “intuition” that the semantic content of 5.1 alone provides a complete and accurate description of John’s cognitive state dissolves when we think about the interpretive process required for the hearer to understand what information the speaker is trying to convey. The assumption that 5.1 provides an unequivocal, transparent, and immediate representation of John’s cognitive state overlooks this interpretive process. It also overlooks the fact that the interpretive process is triggered—in the case of a hearer who knows that “Hesperus” and “Phosphorus” both name the same planet—by the hearer’s apprehension of the fact that something contradictory has been asserted, namely, that John does and does not believe of a certain planet that it is luminous. The standard argument that the semantics of *believes that* sentences requires the incorporation of the believer’s “ways of thinking” about relevant objects—if not in the surface of the sentence then in its logical form—rests therefore on a dubious intuition about the semantic content of sentences like 5.1. This intuition relies on the artificiality of failure of substitutivity puzzle contexts and fades when the allegedly canonical descriptions are imagined in actual descriptive scenarios.

What this suggests is exactly the *reverse* of accounts that appeal to “Gricean” or pragmatic considerations to explain the “liberal” descriptive practices that don’t conform to the expectations of the EFA. Those accounts assume that “distortions” are imposed by ordinary psychological describers because of the limitations of their language, their impoverished epistemological perspective, the peculiarities of their speaking context, the knowledge they share with their audience, and various idiosyncratic communicative intentions they may have. Explanation of these sources of distortion is considered to fall within the realm of pragmatics. But as I have indicated earlier, exactly the same sorts of pragmatic reasoning must be brought in by the interpreter of a belief description in order to understand allegedly “canonical” forms of report, like 5.1. The semantic content of the report by itself does not reveal to the audience whether the speaker is attempting to present John’s belief in a way that reflects its role in his cognitive architecture, or whether it is the *speaker* who is confused about the common reference of the names “Hesperus” and “Phosphorus,” or whether something else is going on. The hearer must rely on contextual clues to figure this out. There is no default assumption—absent some peculiar feature of a particular context—that the belief description provides an isomorphic reflection of a piece of John’s mind. When we understand this, we understand that our judgments about the truth values of 5.1 through 5.3, respectively, should not merely reflect what we think is “really going on” in the mind of the subject of the little cognitive scenario we have been presented

with. We must also imagine how these descriptions would be understood in an actual context of use. By reflecting on such contexts, we realize that 5.1 is not a straightforwardly transparent presentation of the contents of John's mind.

Once we reject the EFA, we are free to formulate a semantics according to which 5.1 really is the contradiction it appears to be. We can still understand why a speaker might have reason to utter 5.1 in order to make a point, specifically, the point John does not understand that "Hesperus" and "Phosphorus" are coreferential and has been led by his ignorance into contradiction. An utterance of this contradiction forces the hearer into an interpretive dilemma that is resolved once the hearer realizes that John's error lies in his failure to understand that these two terms co-refer. The idea that there is a "puzzle" is motivated by the assumption that 5.1 must be true. But we can reject the claim that 5.1 is true while nonetheless accepting that a speaker has reason to utter it.

CONCLUSION

To summarize: I have argued that the puzzles that lead to the complicating of the semantics for *believes that* sentences arise because of the EFA and CFA. I have tried to show that these two assumptions are in turn a product of the roles philosophers have traditionally assigned to semantics—semantics as tantamount to philosophical analysis and semantics as a theory of thought content. Because of the influence of these two assumptions, various substitutivity puzzles and belief puzzles that rely on them have dominated the discussion of *believes that* sentence semantics. The end result has been an overburdening of the semantics for these sentences.

As I have shown, if we abandon the EFA and the CFA, we are free to assume a more straightforward and minimalist semantic strategy (e.g., one that allows for substitution of coreferential terms into belief contexts). Puzzles that have plagued discussions of belief description semantics lose their force. I have argued for rejection of the EFA on the grounds that it represents an unwarranted bias toward one use of belief descriptions sentences while ignoring others. I have argued that we should reject the CFA because it overlooks the other forms of belief description that are used in puzzling and nonpuzzling situations to provide complete descriptions of rational agents. The ubiquity of these historically marginalized belief description practices further shows that the force of the EFA and the CFA has derived not from considerations about semantics per se but rather from considerations about the role semantics should play in other philosophical endeavors.

ACKNOWLEDGMENTS. I wish to thank Joseph Almog, Bill Demopoulos, Joshua Flaherty, Stavroula Glezacos, Bill Harper, David Kaplan, Paolo Leonardi, Peter Ludlow, and Gurpreet Rattan for helpful comments on various drafts of this essay. Special thanks also go to audiences at the University of Maryland, the University of California–Berkeley, and Western Washington University for helpful feedback on talks relating to this material.

CHAPTER 14

What Is *De Re* Belief?

Robert Stalnaker

There are two classic papers about belief that set the stage for most of the subsequent discussion: W. V. Quine's "Quantifiers and Propositional Attitudes," first published in 1956, and David Kaplan's "Quantifying In," published twelve years later. Kaplan's paper made explicit some of what was implicit in Quine's paper and significantly modified the analysis (not to mention providing a rich trove of illuminating asides). I had always thought of "Quantifying In" as a transitional paper, partly looking back to a sententialist/representationalist conception of belief, and partly looking forward to a more externalist conception that is implicit in the direct reference theories later developed by Saul Kripke, Keith Donnellan, and Kaplan himself in his work on demonstratives. But a Fregean, representationalist strain has persisted in Kaplan's later work. In his APA presidential address on *de re* belief, and in later revisions of it, he argued for the thesis that "there is no natural, primitive and pure, *de re* belief. Rather, all belief regarding objects proceeds by way of representation."¹ This is a somewhat dark saying: In a sense, it is uncontroversial that all belief regarding anything proceeds by way of representation, since belief states are representational states. But the slogan points to two contrasts—two ways that belief *de re* is supposed to be indirect in a distinctive way. First, there is the obvious contrast between belief *de re* and belief *de dicto*—beliefs involving objects, and beliefs involving only properties, relations, and general concepts. Second, there is a contrast between cognitive/epistemic modality and metaphysical/temporal modality. Kaplan suggests that metaphysical *de re* modality may be direct in a way that epistemic and cognitive modal notions—*de re* knowledge and belief—cannot be. I think that both contrasts are overdrawn, but I will focus on the first of them in this essay. I agree that, in a sense, the relation between individuals that states of mind are about and the states of mind that are about them is an indirect relation. But the point generalizes. It is

1. Kaplan (manuscript), 2.

not that the relation between individuals and mental contents is indirect. Rather, the point is that the relation between the contents of mental states and the mental states that have the content is indirect in the sense that a content is not a component of an internal state of the thinker but an abstract object used, by the attributor, to characterize an internal state. More generally, we can refer to individuals, kinds, properties, and relations in other possible worlds (for the purpose of characterizing both cognitive and causal/metaphysical modalities) only by using the individuals, kinds, properties, and relations we find in the actual world. So I will agree with Kaplan's thesis about the indirectness of *de re* belief ascriptions, but only as a special case of a claim about belief in general, *de dicto* and *de re*: There is no natural, primitive, and pure *de dicto* belief. Rather, all belief involving objects, properties, relations, and propositions proceeds by way of representation. Propositions are objects that are used by the attributors of propositional attitudes to characterize states of mind.

My main aim in this essay will be to bring out some of the interconnections between the specific issues about the logical form of *de re* belief attributions and the broader issues about intentionality. I will begin by going over some familiar ground, reviewing the Quine/Kaplan analysis, as I understand it, spelling out the problem that originally motivated it, and considering some different assumptions about the objects or contents of belief. I will then look at Kaplan's modifications of Quine's analysis, and at a way that it might be further generalized, in light of its internalist motivation. I will conclude by looking at *de re* belief, and belief generally, from a contrasting externalist perspective.

So what is a *de re* attitude ascription? The short and uncontroversial answer is that it is a description of a person's state of mind in terms of—as a function of—an individual that the person's thought is about. To ascribe a belief *de re*, you identify a believer (say, Ralph), an object (say, Bernard J. Ortcutt), and an attribute (say, being a spy) that the believer attributes to the object. Ralph believes of Ortcutt that he is a spy.

The problem, first raised in Quine's classic article, was this: Belief ascription contexts are paradigm examples of contexts that are *referentially opaque*, or *indirect*. That is, they are contexts in which one cannot freely substitute coreferential singular terms. ("Ralph believes that the vice president is in hiding" might differ in truth value from "Ralph believes that Dick Cheney is in hiding" even though the vice president is Dick Cheney.) Quine had argued that quantification into opaque contexts was unintelligible. The reasoning was something like this: A quantified statement is a statement that says that some predicate is true of everything, or something (in the relevant domain). So to make sense of a quantified statement, we have to understand the open sentence that is in the scope of the quantifier as expressing an attribute or property—something that is true or false of things. Ordinary statements containing names can be understood as predicating something of the thing

named, but the substitution failures that mark opaque contexts indicate that the name is not being used simply to refer, and so the statement as a whole cannot be understood as saying something about the individual that the name refers to. Quine's model for an opaque context is quotation (and he thought that ultimately, all opaque contexts could be explained in terms of quotation). To quantify into a belief context would be like quantifying into quotation.

But—and here is the problem—it seemed clear to Quine that one *can* quantify into belief contexts. A sentence such as

- (1) "Ralph believes someone is a spy"

seems to be ambiguous, with the ambiguity explained by a quantifier scope difference. On the weak reading, it says that Ralph has the general belief that there exist spies, while on the stronger reading, with the quantifier given wide scope, it says that there is some particular person who is believed by Ralph to be a spy. To make sense of this reading, we need to understand "believed by Ralph to be a spy" to express an attribute said to be true of someone (the person Ralph's belief is about). If the existential statement is true, then there will be an individual (Bernard J. Ortcutt, as it happens) it is true of. But what does it mean to say that Ortcutt—the man himself, however he is referred to—is believed by Ralph to be a spy?

That is Quine's problem; what is his solution? The first move—more of a setting up of the problem than a solution to it—was to reject the straightforward quantifier scope account of the ambiguity in (1). What is rejected is the regimentation of the two readings of this sentence in terms of a univocal belief operator, *B*, in the following straightforward way:

- (2) $B(\text{Ralph}, (\exists x)x \text{ is a spy})$
 (3) $(\exists x)B(\text{Ralph}, x \text{ is a spy})$

Since the belief operator in (2) creates an opaque context, one cannot (according to Quine) quantify into it, as is done in (3). So we need to distinguish two different belief operators that differ in their syntax as well as their meaning. What Quine calls *notional* belief is regimented with an operator on propositions, as in (2). What Quine calls *relational* belief is regimented with an operator that has two or more operands (in addition to the believer)—an individual and an *attribute*. Instead of (2) and (3), we use (4) and (5) to represent the distinction between the two interpretations of (1):

- (4) $B_N(\text{Ralph}, (\exists x)x \text{ is a spy})$
 (5) $(\exists x)B_R(\text{Ralph}, x, \lambda y(y \text{ is a spy}))$

The Quinean analysis, unlike the straightforward one, will yield an ambiguity even in the unquantified sentences that say something more specific about Ralph's beliefs. Quine distinguishes the notional or *de dicto* belief ascription that Ralph believes that Ortcutt is a spy from the

relational or *de re* ascription that Ralph believes of Ortcutt that he is a spy:

(6) $B_N(\text{Ralph}, \text{Ortcutt is a spy})$

(7) $B_R(\text{Ralph}, \text{Ortcutt}, \lambda y(y \text{ is a spy}))$

The name "Ortcutt" appears in referential position in (7), but not in (6). That is, in (7), but not (6), the *speaker* is referring to Ortcutt, and describing Ralph's beliefs in terms of him.

As I said, Quine does not regard this regimentation as more than a first step, getting the logical form right at a shallow level of analysis. Two problems remain: First, one needs an explanation (for Quine, an elimination) of the problematic notions of *proposition* and *attribute*. Second, we need an account of the logical relation between the two different concepts of belief—preferably a reduction of one to the other.

In the original article, Quine made only one brief remark, in passing, about the second problem, the relation between notional and relational belief: "The kind of exportation which leads from [(6) to (7)] should doubtless be viewed in general as implicative."² But as Kaplan later brought out, this assumption is enough to motivate a reduction of relational to notional belief. We will look at Kaplan's analysis later.

On the first problem: Quine was of course deeply skeptical of all intentional notions, and while he wanted ultimately to reduce them away, his strategy was to proceed in stages, taking such things as propositions for granted to get the logical form right at a shallow level of analysis and then moving on to analyze them away. "Intensions," he wrote, "are creatures of darkness, and I shall rejoice with the reader when they are exorcised, but first I want to make certain points with the help of them" (Quine [1956], 186). At a later stage, his plan would be to analyze propositional attitudes in terms of relations to sentences, replacing both notional and relational concepts of belief with predicates that apply to linguistic objects—sentences and predicates. Notional believing was to be explained in terms of believing-true a sentence, and relational believing was to be explained in terms of believing-satisfied a certain predicate by an individual.

Quine remarked that the sentential reduction does not require one to assume that the subject speaks the language in which the sentence occurs. The idea is that we, the ascribers of beliefs, use the sentences of our own language to describe the mental states of others who may speak other languages, or none at all. Quine said, rather cavalierly, "We may treat a mouse's fear of a cat as his fearing true a certain English sentence. This is

2. Quine (1956), 188. Quine later qualified this remark because of the possibility of a notional belief involving a nonreferring name. One cannot infer from "Adam believes that Santa Claus brings presents on Christmas" that Adam believes of Santa Claus that he brings presents on Christmas.

unnatural without being therefore wrong. It is a little like describing a prehistoric ocean current as clockwise" (192). This remark points to an externalist version of a sententialist strategy, one that is closer in spirit to the antirepresentationalist conception of intentionality. We will look back, briefly, to the linguistic reduction later, but for the moment (and, if we are wise, forever) we will stay with the creatures of darkness.

We have been talking about *attributes* and *propositions*. At this point, we want to assume as little as possible about the nature of these things. We assume only that an attribute is something true of an individual. About propositions, we assume only that they are the objects, or contents of (notional) belief—what is expressed in a statement, and what is denoted by a that-clause. A sentence like "Ralph believes that some pigs can fly" involves two propositions, one about what Ralph believes, and one about pigs. The first is a function of the second. So "Ralph believes" is a *propositional operator*—its meaning takes one proposition (that some pigs can fly) into another (that Ralph believes that some pigs can fly).

Just one further bit of jargon: A *propositional function* is a function from individuals to propositions.

Now whatever propositions are, any propositional operator will also be an operator taking propositional functions into propositional functions. If $f(x)$ is a propositional function, taking an individual x into a proposition, then "Ralph believes $f(x)$ " will also be a propositional function—one that takes the individual x into the proposition that says that Ralph believes the proposition that is the value of f for the argument x . So much is clear and, at least if one assumes there are propositions, uncontroversial; this observation assumes nothing about what propositions are. But suppose we make one natural assumption about the relation between propositional functions and attributes, which I will label assumption (S):

(S) Any attribute determines a unique propositional function that takes an individual into the proposition that attributes the attribute to the individual, the proposition that is true if and only if the individual possesses the attribute.

Propositions in general say how things are—what the world is like. It seems reasonable to assume that the simplest and most basic propositions will say something about a particular individual; they will say something about how things are with respect to that individual by ascribing some attribute to it.

Assumption (S) still makes no specific commitment as to what a proposition is. A Russellian conception of proposition will conform to (S). On this conception, propositions are complexes—ordered sequences with a recursive structure that allows ordered sequences as constituents, and with individuals, properties, and relations as the ultimate constituents. But a contrasting conception according to which propositions are

individuated coarsely, and represented by the sets of possible worlds in which they are true, will also conform to (S). What (S) does involve is a commitment to *singular propositions*, which can be defined as propositions that are the values of the propositional functions that are defined in terms of attributes.

The Fregean conception of proposition (or Thought), however, will not conform to (S). For Frege, a proposition is a kind of sense, and Fregeans usually assume that Thoughts can be represented as structured complexes whose ultimate constituents are the senses of names and predicate expressions. A sense of a singular term is some kind of abstract object that determines an object as its referent but that allows that there may be different senses determining the same referent. On a Fregean conception, the proposition that Ortcutt is a spy might be an ordered pair consisting of the sense of the name "Ortcutt" and the sense of the predicate "is a spy." There will be no unique proposition determined by the *person* Ortcutt and the attribute of being a spy, since there may be different senses that determine Ortcutt as their referent.

For Russell, singular propositions had particular objects as constituents, and Frege apparently found this to be an absurd consequence. In a famous, often-quoted exchange of letters, Frege wrote, "Mont Blanc with its snowfields is not itself a component part of the thought that Mont Blanc is more than 4000 meters high." The idea that a concrete object, as big as a mountain, should be a *constituent* of the content of a thought may seem strange, but Russell bites the bullet: "I believe that in spite of all its snowfields Mont Blanc itself is a component part of what is actually asserted in the proposition, Mont Blanc is more than 4000 meters high."³ Should this be a bullet that is hard to bite? There is a temptation to think of propositions as internal to the mind, and if one does, it will seem absurd to allow that they might contain mountains as parts. (Putnam taught us that meanings ain't in the head, but it was clear all along that there ain't any mountains in there.) But for Frege and Russell both, propositions or thoughts are abstract objects used to characterize states of mind and are not things to be found in the head. The term *content* for proposition helps to promote a false picture. (A representation is a vehicle, and its propositional content is riding inside.) This is a picture that both Russellians and Fregeans should reject.

If one accepts (S), one can make sense of the simple scope analysis of *de re* belief, but any conception of the object of (*de dicto*, or notional) belief that rejects (S) must follow Quine in rejecting the simple analysis. The Fregean conception of proposition does, however, point to a way of understanding Quine's *de re* or relational sense of belief, a way of reducing it to notional or *de dicto* belief. Kaplan, in "Quantifying In," spelled

3. Frege to Russell, November 1904, and Russell to Frege, December 1904. The letters are in Frege (1980), 169.

out the analysis, which he argued was implicit in Quine's discussion, and defended a modification of it. Let me introduce a bit of notation in order to spell out a version of the Quine/Kaplan analysis that presupposes a Fregean framework:

Let "s" be a variable ranging over senses, and let d be a function taking any sense into the corresponding denotation. So if the value of s is the sense of the name "David Kaplan," $d(s)$ will be the person, David Kaplan. Assume that simple Fregean propositions are ordered pairs consisting of a singular term sense and an attribute.⁴ So if O is the sense of the name "Ortcutt," the Thought that Ortcutt is a spy is the pair $\langle O, \text{spyhood} \rangle$. The analysis of Quine's relational sense of belief,

$$(7) B_R(\text{Ralph}, \text{Ortcutt}, \lambda y(y \text{ is a spy}))$$

will then be as follows:

$$(8) (\exists s)(B_N(\text{Ralph}, \langle s, \text{spyhood} \rangle) \ \& \ d(s) = \text{Ortcutt})$$

That is, the claim that (7) makes is that there is some sense or other that determines the person Ortcutt as its denotation, and that occurs in a constituent of a belief of Ralph's, along with the attribute spyhood. So the relational belief ascription is made true by a belief of the ordinary notional sort, but the relational belief attribution does not say what specific belief of that kind it is that Ralph has. Rather, the relational belief ascription makes an existential claim, the claim that Ralph has some belief or other that meets a certain condition, and the condition that it must meet is an *external* condition, which is to say that it is a condition that is independent of the mental state of the believer. One could imagine a counterfactual situation in which Ralph's *notional* beliefs were exactly as they actually are, but in which he did not believe of Ortcutt that he was a spy. This would be a counterfactual situation in which the relevant sense that in fact picks out Ortcutt instead picked out someone else.⁵ This analysis reconciles an internalist conception of belief with an externalist account of *de re* belief ascription by interpreting the belief ascription as an extrinsic and indirect characterization of a state of mind. The analysis gives concrete form to Kaplan's later claim that "there is no natural, primitive and pure, *de re* belief."

The Quine/Kaplan analysis recognizes a way in which the characterization of intentional states of mind is indirect and extrinsic, but it retains

4. For simplicity, I will ignore, for the moment, the fact that it should really be the sense of a predicate, rather than an attribute or property, that is a constituent of a Fregean Thought. Predicate senses are potentially individuated more finely than the attributes they determine, just as singular term senses are individuated more finely than by their referents.

5. Not every neo-Fregean will accept that it is possible for a sense to determine a referent different from the referent that it actually determines. Some (such as John McDowell) would argue that senses may be object-dependent. But to the extent that Fregeans are motivated by the internalist picture, they should admit this possibility.

the internalist idea that the relation between thinkers and the propositional contents of their thoughts is an intimate and direct one, locating the indirectness in the relation between the characterization of the state of mind and the propositional content of the state characterized. What is retained is the idea that beneath the extrinsic characterization is an object of belief that is *grasped* by the believer, a content that is, in a sense, wholly internal to the mind; it is the belief as it is in itself. A contrasting response to the phenomenon of *de re* belief ascription is to see it as a particularly striking manifestation of a more general feature that characterizes all intentionality. In general, to characterize a state of mind as representational—to see it as representing the world to be a certain way—is to describe (from the ascriber's perspective) a way the world might be, and then to use that characterization to say how the thinker in question takes it to be. The contrasting response seems to me the right one, but I will approach it by developing further the Quine/Kaplan picture.

Quine's account of relational belief, and Kaplan's reduction, on Quine's behalf, of relational to notional belief succeeded in showing that apparent quantification into belief contexts could be reconciled with Quinean principles, but Kaplan argued that the analysis had intuitively unacceptable consequences and so needed modification. The problem is that the phenomena that originally needed explanation suggested a contrast between indirect, purely descriptive propositions, such as, for example,

- (9) The person who will be the Democratic candidate for president of the United States in 2008 is a woman

and beliefs about particular individuals, such as

- (10) Hillary Clinton is a woman.

But as Kaplan brought out, the Quinean analysis implied that if I believe (9), then if it turns out that the Democrats in fact are going to nominate Barack Obama in 2008, it will be true to say of me that I believed Barack Obama to be a woman, and this is not a plausible result. In response to this kind of problem, Kaplan proposed to keep the basic structure of his Quinean analysis of relational belief but to add some restrictions that eliminated such counterexamples. The contents of thought remained essentially descriptive, and it was still held that a *de re* belief ascription made an indirect, existential claim about the content of belief, but the analysis was strengthened so that such ascriptions made a claim about the *kind* of descriptive sense that must be a constituent of the Thought. To believe Ortcutt to be a spy, you must have a descriptive belief involving an Ortcutt-determining sense, but it was also required that this sense be one that presents Ortcutt in a particular way: One must have a rich conception, a vivid mental name, that was appropriately related to Ortcutt. One must be, to use Kaplan's term, *en rapport* with Ortcutt, or to

use the more traditional term, one must be *acquainted* with Ortcutt in a way that Kaplan's proposal spelled out. The simple requirement that the sense denote him was replaced by a stronger relation between the sense, or mental name, the believer, and the denotation. But there is something ad hoc about this move. The idea that one must be *acquainted* with an object to have a *de re* belief about it was part of the contrasting picture of *de re* belief—belief in a singular proposition. The original motivation for an acquaintance requirement was that acquaintance was necessary to grasp the singular proposition. But that motivation drops away when the singular propositions are no longer in the picture. And while Kaplan's replacement for the acquaintance relation does eliminate the obvious counterexamples that it was designed to eliminate, it seems too restrictive. Intuitively, *de re* belief attributions are not so demanding. I don't have to have a very rich concept of Cicero, or Uzbekistan, or the prophet Isaiah in order to have beliefs about them that are appropriately ascribed *de re*. Most of the examples that Saul Kripke used in *Naming and Necessity* to argue against a description theory of names will also provide cases where *de re* belief ascription is reasonable, even though the believer has a very limited conception of the thing or person the belief is about.

Even if the Quine/Kaplan analysis of *de re* belief ascriptions were to succeed in explaining them without putting individuals into the contents of belief, it may need to be extended in order to give a fully internalist account of content. In sketching Kaplan's account, we took the attribute, or property (spyhood), to be a constituent of the proposition. It may be easier for an internalist to think of properties, as contrasted with mountains, as constituents of propositions, but strictly they should not be there either. If you think of Fregean thoughts as somehow in the head, you shouldn't expect to find such things as being a spy, being made of aluminum, being taller than the average adult male in Uzbekistan in them, any more than one should expect to find Mont Blanc. And while identity confusions with properties are not so easy to come by, there is no reason, in principle, why the same attribute cannot be presented to us in different ways. Externalism, or anti-individualism, about mental and linguistic content began with singular reference but was generalized early on by Hilary Putnam and Tyler Burge (as well as Kripke), and the standard internalist response to the phenomena highlighted by Putnam and Burge was analogous to the Quine/Kaplan strategy for analyzing *de re* belief ascription and can be seen as a kind of generalization of it. Just as a *de re* belief ascription can be understood as a claim that the believer has an unspecified belief of a certain kind, so externalist belief ascriptions generally might be understood as claims that the believer has a belief with an internal content that meets a certain general condition but that is not fully specified. An analysis of wide-content belief in terms of a narrow-content concept of belief might fit the following form:

(11) x believes that $P =_{df} (\exists Q)(x \text{ believes}^* Q \ \& \ R(Q, \text{that } P))$

This is the abstract form of an analysis that can be seen to parallel the Quinean analysis of relational belief in terms of notional belief: Ordinary belief in an externally individuated proposition (analogous to relational belief) is defined in terms of an existential generalization over an internally individuated kind of belief (analogous to notional belief), where an extrinsic condition—a relation between the narrow and wide content (analogous to the denotation condition, in the unreconstructed Quine analysis, and to the “en rapport” relation in Kaplan’s revision of it).⁶

But this kind of generalization and extension of the Quine/Kaplan analysis changes the game in a significant way. The analysis of *de re* belief in terms of *de dicto* belief is the explanation of one special case in terms of the ordinary concept of belief, while the generalization analyzes belief, in the ordinary sense, in general, in terms of a hypothesized theoretical notion that remains to be specified, and to be related to ordinary belief. What motivated the Quine/Kaplan analysis was the attempt to reconcile the phenomena pointing to an extrinsic individuation of mental states with an internalist conception of belief itself. So long as the evidence of extrinsic individuation was isolated to a special case, one could analyze it away in terms of the general case, but when extrinsic individuation is seen to be the general case of ordinary attribution of content, then it is no longer clear that there are any conceptual resources that are both intentional and internalist. The internalist assumes that even if we normally characterize mental states only indirectly, there must be the *real* content of the mental state that we describing only indirectly. John McDowell gives an apt description of this internalist picture, which he thinks is “a suspect conception of how thought relates to reality”:

The Fregean view would have to seek its support in the idea that thought relates to objects with an essential indirectness: by way of a blueprint or specification which, if formulated, would be expressed in purely general terms. . . . When we mention an object in describing a thought we are giving only an extrinsic characterization of the thought (since the mention of the object takes us outside the subject’s mind); but there must be an intrinsic

6. Hartry Field proposed an analysis fitting this form in Field (1978). Field’s version was a sententialist account: The “believe” relation was a relation between a believer and a mental sentence, and the “R” relation was a semantic relation between the sentence and its meaning or content. The idea of Field’s analysis was to factor the mental-intentional belief relation into a psychological (functional) but nonintentional relation between a person and a (mental) sentence and a semantic but nonpsychological relation between the sentence and its content. The idea was to locate the problem of intentionality as a problem within semantics. But an internalist project that postulates some notion of narrow content (an abstract object apt for describing the purely internal component of an intentional state) might take the same form. The form of analysis brings out the challenges that such an internalist needs to meet: What is needed to give substance to the abstract analysis is an account of the internal belief relation, of the objects that are narrow contents, and of their relation to ordinary, externally individuated propositions.

characterization available (one which does not take us outside the subject's mind), and that characterization would have succeeded in specifying the essential core of the thought even if extra-mental reality had not obliged by containing the object.⁷

The alternative to this internalist picture is to think of the extrinsic characterization as essential to the idea of intentionality. To think of something as a believer is to think of it as having a point of view on the world, and to ascribe beliefs to it is to attempt to describe that point of view. The way we do this is to say what the world is like according to them, which is to say, in our own terms, using the resources we find in the world as *we* take it to be, what the world is like according to them. We (the attributors) refer to things, kinds of things, properties, relations, and events, and we use those materials to describe the world as someone (the person to whom we are attributing mental states) takes it to be. It is not that this is a rough, approximate, or indirect way of characterizing the contents of a person's mind, contents that might, if we could only see things as they do, be described more directly. It is not that if we were to open up the person's mind and stare it in the face, we would find the content as it is in itself. There is only a brain in the head—no pieces of information, no contents, wide or narrow. Propositions, whether they are Fregean, Russellian, or analyzed (as I would prefer) in terms of possible worlds that are the way the person takes the world to be, are abstract objects apt for describing people's states of mind in the way that numbers are apt for describing quantitative physical properties such as weights and lengths.

If we think of intentionality this way, then we get a different perspective on what is distinctive about *de re* belief and *de re* belief ascription. We have no reason to be suspicious of singular propositions, and of the idea that a *de re* belief ascription is the ascription of a particular belief—a belief with a propositional content that is picked out as a function of an individual. But what is it, on this picture, to believe a singular proposition—a proposition about an individual as it is in itself? One can correctly and aptly ascribe a *de re* belief whenever one can give a correct and unambiguous characterization (in a given context) of the way the world is according to the believer by referring to a certain individual and saying that, in the world according to the believer, that individual is a certain way.

There is, of course, a daunting general problem of saying what it is for a way the world might be (or a set of such ways) to represent the way the world is according to some person. The problem of intentionality is the problem of saying how believers must be related to the objects, properties, and relations that are used to characterize the world according to

7. McDowell (1977), 174–175.

them in order for the characterization to be correct. But what is required for a believer to have a belief about a particular individual is not some very strong acquaintance relation with the individual, or some rich and detailed conception of it. This is neither necessary nor sufficient. As the familiar puzzle cases (the Babylonians and Hesperus/Phosphorus, Ralph and Ortcutt, Pierre and London/Londre, Lois and Superman/Clark Kent, etc.) bring out, one may be strongly acquainted with some object or person in different ways. In such cases, one cannot (at least without some special contextual clues) aptly and unambiguously characterize the world as it is according to that person as a function of the individual. It is clear from Quine's story, for example, that in the world according to Ralph, there are two different persons (the man on the beach and the man in the brown hat), one a spy and one not, both of whom are related to Ralph (in the world according to him) in ways that Bernard J. Ortcutt is actually related to Ralph. Which one of these men in the world according to Ralph is the real Ortcutt? No plausible solution to the problem of intentionality will yield a determinate answer to this question, which means that one cannot unambiguously characterize the world according to Ralph as a function of Ortcutt. On the other hand, there are cases in which one's conception of someone or something is thin and mainly inaccurate but is nevertheless a conception that uniquely connects the believer to a particular individual, at least in the relevant context. In this kind of case, a *de re* belief ascription, a belief ascription with a singular proposition as the content ascribed, may be apt, and correct, despite the believer's impoverished conception. Even if I know little about Uzbekistan, and even if some of my few beliefs about it are wrong, it is still Uzbekistan that I believe to be a former Soviet republic; in the world according to me, it is the actual Uzbekistan that used to be a Soviet republic.

As Kaplan and others have noted, there are cases of *de re* belief ascription where the believer clearly lacks anything close to the kind of acquaintance relation that he takes to be required for a normal *de re* belief ascription. He labeled such cases "pseudo-*de re*" and suggested that they needed special treatment. But from the perspective I am promoting, there need be nothing "pseudo" about such cases. When someone characterizes the world according to a believer, he or she is selecting a kind of possible state of the world from among a contextually determined set of possibilities, the possibilities that might, for all the attributor is presupposing, be a way the world might be according to the believer. Whether an attributor can make this kind of selection by referring to an individual will depend on the set of possibilities from which the selection is made, and when the context is highly constrained, it may be possible to describe the world according to a believer as a function of an individual even when the believer does not, in the ordinary sense, know who the individual is. I will conclude by looking at one of Kaplan's examples of the allegedly pseudo-*de re* from this point of view.

David Kaplan is late for a talk that he is scheduled to give, but he has called ahead to say that he will be late, and it has been announced to the audience that the speaker has been delayed. As Kaplan and his wife make their way to the site of the talk, he says to her, "Don't worry, they know I am going to be late." What they know is that the speaker will be late, but (let us suppose) they don't know who the speaker is. Yet Kaplan seems to have ascribed to the audience *de re* knowledge of Kaplan that he will be late.

What is presupposed, by Kaplan in his conversation with his wife, is that the audience knows that there is a (unique) speaker scheduled to give a talk at a certain time. That is, it is common ground (between Kaplan and his wife in the context in which he made his remark) that in each of the possible situations compatible with what the audience knew, there is a unique salient person who is scheduled to give a talk. Since Kaplan and his wife know that the speaker in question is Kaplan, in their context of attribution, they can use possible worlds in which the speaker is Kaplan to represent the possibilities that the audience's knowledge is distinguishing between. And if those are the possibilities in question, then the appropriate subset of them (the ones in which the speaker will be late) can be selected as a function of Kaplan—the person himself. In that context, it is correct to attribute to the audience belief in the singular proposition that Kaplan will be late. The reference to Kaplan is made by the attributor and is extrinsic in the sense that the attribution does not locate the referent in the mind of the members of the audience to whom knowledge is attributed. The claim is that this is the way things are in the general case when an attributor refers to a proposition in order to describe a state of mind.

REFERENCES

- Field, Hartry. (1978). "Mental Representation." *Erkenntnis* 13, 78–112.
- Frege, Gottlob. (1980). *Gottlob Frege: Philosophical and Mathematical Correspondence*. Edited by G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart. Abridged from the German edition by B. McGinness. Translated by H. Kaal. Chicago: University of Chicago Press.
- Kaplan, David. (1968). "Quantifying In." *Synthese* 19, 178–214.
- Kaplan, David. (manuscript). "De Re Belief."
- McDowell, John. (1977). "On the Sense and Reference of a Proper Name." *Mind*, n.s., 86, 159–185.
- Quine, W. V. (1956). "Quantifiers and Propositional Attitudes." *Journal of Philosophy* 53, 177–187. Reprinted in Quine, *Ways of Paradox and Other Essays*, 183–194. New York: Random House, 1966. Page references are to the latter.

CHAPTER 15

Five Theses on *De Re* States and Attitudes

Tyler Burge

I shall propose five theses on *de re* states and attitudes.* To be a *de re* state or attitude is to bear a peculiarly direct epistemic and representational relation to a particular referent in perception or thought. I will not dress this bare statement here. The fifth thesis tries to be less coarse. The first four explicate and restrict context-bound, singular, empirical representation, which constitutes a significant and central type of *de re* state or attitude.

The five theses are developed against a background rejection of Russell's notion acquaintance, a supposed perspective-free mental relation to an object. I regard Russell's view of reference as psychologically and epistemically naive. Analogs of the view have some recent advocates—both in naive realism about perception and in direct-reference views about language transferred whole to perception and thought. I regard such views as both empirically and conceptually untenable. I take Russell's view and its successors to be useful mainly as a foil or limiting position.

The theses are also developed against a background rejection of the view, often associated with Kant (mistakenly, I believe), that to perceive

* The first four sections of this essay are based on sections IV through VI of "Descartes and Anti-Individualism: Reply to Normore," in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, ed. Martin Hahn and Bjorn Ramberg (Cambridge, Mass.: MIT Press, 2003). The present essay's counterpart passages involve very considerable development, correction, and elaboration of the earlier work. Here I also abstract from any supposed relation to Descartes. The arguments for the second and third theses are new, and a great deal of the discussion of the second thesis is new. I have replaced the term "formally general," which occurred in the earlier paper, by "semantically general" here. The present section IV on apriority is substantially rewritten. Section V is entirely new. The new parts of the essay were mostly written in 2003–2004. Publication of the essay was unfortunately delayed for some years by differences among other parties over the form of the volume. Substantial expositional revisions in section II were entered in 2007. The essay has benefited from comments at Syracuse University, Princeton University, and UCLA, especially a comment by Daniel Nolan. I have also benefited from discussion with Louis DeRosset and Luca Struble.

a physical object, an individual must apply a battery of conceptual or linguistic resources. Examples of resources that perception of objects is supposed to depend upon are a conception of causal relations; an ability to locate oneself in an objective spatial framework; quantification, cross-reference, and identity; and so on.

Against Russell's view, I believe that perception and thought are fundamentally and ineliminably perspectival. Against the neo-Kantian view, I believe that it is clear—scientifically established—that perception, even of bodies, need not rely on conception, certainly not on any sorts of conception postulated in the neo-Kantian tradition. I will not discuss these alternative views here. I mention them only for orientation.

The first thesis formulates the perspectival nature of representation. The second outlines attributional resources necessary to perceptually based representation. These resources omit Russellian acquaintance but develop Russell's insight that singular representation begins at a primitive, preconceptual level. The third thesis holds that some of these attributional resources must apply veridically if perceptually based singular reference is to occur. The fourth sketches how these resources provide a basis for apriori knowledge, although they are much less rich than those postulated by the neo-Kantian views. The fifth outlines a nonempiricist conception of *de re* states and attitudes that builds on the empirical cases that dominate sections II and III. All the theses except the first are proposed in a conjectural spirit.

I start by saying a little about representation. Examples of representations are perceptual contents, concepts, representational thought contents, words, numerals, recordings, musical scores, photographs, diagrams, mimetic paintings. I take *mental* representations—including perceptual contents, concepts, and representational contents of thought—to be the basic sorts.¹ I shall concentrate on them. I assume a distinction between perception and propositional thought, and a companion distinction between certain components of their representational contents—perceptual attributives and concepts. I note differences as I go, but much of what I say applies to both perception and thought.

The title alludes to non-propositional representational *states*, particularly perceptual states, and propositional *attitudes*. I intend this nomenclature to be broad and fluid. My term “non-propositional representational state” is meant as a catch-all—to include perceptual events, perceptual states, perceptual capacities, perceptual memories, perceptual anticipations,

1. I am taking for granted here a distinction between perceptual representation and conceptual representation, as I did in “Cartesian Error and the Objectivity of Perception,” in *Subject, Thought, and Context*, ed. John McDowell and Philip Pettit (New York: Oxford University Press, 1986). Also see “Perceptual Entitlement,” *Philosophy and Phenomenological Research* 67 (2003): 503–548. I shall discuss the distinction in future work.

Throughout the essay, I use italics for emphasis or for designating foreign words, and underlining to denote representational contents.

perceptual imaginings, perception-dependent intermodal states, and perception-guided actional states. My formulations of the theses for perception, perceptual memory, and so on, should be understood to cover all empirical states of these kinds that are representationally directed toward the world. The restriction is only that the “state” is to be representational but not propositionally structured. My term “propositional attitude” is also meant as a catch-all. It includes propositional states, events (including acts), and capacities—broadly speaking, propositional thought.

The ontology of mental representation is largely unimportant here. With caveats noted in section II, I take mental representations to be abstract representational kinds, *not* particulars. Thus it is not assumed that representations, or representational contents, are “mental objects,” or token entities in individual minds. Individuals can share representations. Representations are ways of thinking or perceiving. So they have “intentionality” or representationality. I do not distinguish between mental representations and mental representational contents. I assume that individuals have representational *states* and *capacities*, undergo representational *events*, and engage in representational *acts*. Representations (representational contents) *mark* or *help type-identify* such states, capacities, events, or acts. I leave open whether *instances* of such contents, representation *tokens*, are always present in mental states or events in any sense beyond the fact that state or event instances which the representations mark are attributable to individuals. So I leave open whether conceptual representations are always associated with a separately specifiable language of thought whose “words” are tokens of the conceptual representations. (I do hold that such a language’s syntactical tokens must be partly type-identified by their representational content.) I also leave open how representations and representational states and events relate to neural states.

The reason why I regard concepts, perceptions, thought contents—indeed all mental representations—as abstractions, not tokens in minds and not “mental objects,” is that I believe that fundamental explanatory enterprises invoke in-principle shareable contents. The relevant fundamental explanatory enterprises are psychological explanation, accounts of reference, and accounts of warrant and knowledge. The abstract contents, as aspects of kinds, are essential for the explanatory and evaluative aims of these enterprises. These enterprises do depend on reference to the contents and to the states and events in the individuals. The contents mark the states and events. And the states and events are tokened or instantiated in individual persons or animals. But these enterprises do not always *clearly* depend on reference to instances of contents over and above the states and events. I have no doubt that some representations—conscious perceptions, for example—do have instances or vehicles “in” individual minds. But I am not committed to the view that all do. Thus many standing states need characterization in terms of representational contents. But whether there is a further instantiation of the content

in all such cases is, I think, a less empirically and explanatorily established matter. I think reference to the abstract representational contents is well established in empirical and normative enterprises.

The key points about representations—or representational contents—for our purposes concern their explanatory roles. There are three principal roles. First, representations are *about*, purportedly about, what is represented. A mental representation functions to represent. Some representations fail to represent, but they still function to represent. Mental representation helps constitute the representational perspective of an individual on a subject matter. Second, representations *mark* or *help type-identify* an individual's representational states, capacities, events, acts. Thus they are aspects of kinds of psychological states that are referred to in psychological explanation. Third, representations serve as ground for the application of representational and epistemic norms. As regards representational norms, an individual's representational states are evaluated by reference to whether representations are correct, true, or veridical. As regards epistemic norms, an individual's representational states are evaluated for warrant, rationality, and other types of cognitive “doing-well”—for how the use of representations meets certain standards, given the individual's perspectival and cognitive limitations. Reference to mental representations is well established in both scientific psychology and common sense.

I

The first thesis is that *mental representation is always representation-as*. The thesis rules out any view that maintains that one perceives, conceives, or thinks about objects, properties, or relations without doing so in any particular way that constitutes some perspective on them.² Any

2. This thesis is similar to David Kaplan's slogan “No mentation without representation.” I do not know whether the slogan is supposed to entail that all mental phenomena are representational. The telegraphic term “without” allows various logical forms, and “mentation” is unspecific as between “mental phenomena” and “mental functioning processes.” So I do not know whether I accept the slogan. If it entails that all mental phenomena are representational, then I do not accept at least one reading of it. (Here I am using *my* notion of representation, which requires a degree of objectification, at least the simple sort involved in genuine perception. Mere functionally useful correlation does not suffice for representation in my sense.) I believe that there are qualitative mental phenomena that are not in themselves representational. For example, there is disfunctional qualitative “noise” in psychological systems. Moreover, I am doubtful that all mental (phenomenal) features of representational states are in themselves representational.

I have held the thesis that I state in the text for as long as I can remember. Whether or not my thesis and Kaplan's slogan use the same notion of representation, they both entail acknowledging that Frege's problem can arise for any particular position in a representational content: No matter how an entity is referred to, denoted, or indicated in thought or

view that rejects this thesis fails to accord with fundamental features of perception and thought.

The thesis is to be taken in this specific sense: *Every purported application, reference, and attribution in every content position in all thought and perception is perspectival and is carried through in a perspectival way: it is marked by some representational content, which constitutes a perspectival way of thinking or perceiving.*

I use “represent as” to entail “represent in a perspectival way” or “represent via representational content.” And I take representing in a perspectival way to be equivalent to representing’s having a mode of representation and to representing with representational content.

In perception we represent only through abilities that provide partial, incomplete, usually fallible perspectives on an actual or purported subject matter. Here the notion of perspective is concrete, commonly spatial-directional, sometimes phenomenological. One can have different perceptual representations from different perceptual perspectives on the same property, even representing it as the same property. This is the essence of perceptual constancy. Perceptual constancy is the ability to perceive the same object or property *as* the same object or property even though the perceptual mode of presentation, the perspective on the object or property, varies. The difference in perspective can derive from spatial, temporal, or phenomenological differences. One can also represent the same property in different sense modalities. These also commonly constitute different perspectives.

Parallel points apply to conception. With respect to conception in general, my term “perspectival” is more abstract. I am not specifically concerned with an individual’s particular spatial or phenomenal angle on a subject matter, as one is in perception and egocentrically based spatial, empirical thought. Nothing so concrete is at issue in many cases of conception.

I take any representational content to *constitute* a perspective inasmuch as it is one of many possible ways of representing the same entity—one of many possible representational modes, representational contents. Our conceptual perspectives are not exclusive. They are to be distinguished from the entity itself. They constitute one of many ways of representing the same entity. And, normally, such ways can be correctly or incorrectly applied. Their application is fallible. In these abstract senses, they are perspectives on any entity that they succeed in representing. When representation fails, it still constitutes a perspective. The representer

perception, the same entity can be referred to, denoted, or indicated from a different perspective—marked by a different mental representation, a different representational content. Uniting the different perspectives constitutes a possible achievement, insight, acquisition of information, or realization of cognitive value. Note that for me singular context-bound applications (and pure demonstratives like *that*) count as representations-as. Subsequent theses place more restrictive conditions on singular representation.

can use different representational contents, even though there is no successful *representatum*. One can even represent, with different representational contents, what is purportedly the same entity, even though there is none.

Representation in both perception and propositional thought is type-identified to reflect representational abilities. It is not type-identified purely with regard to what its referent is. I write of such abilities as being perspectival on represented particulars, properties, relations, and kinds. We cannot perceive or conceive of anything without doing so in some way. The perspective or representational content is always one of many that could actually or purportedly apply to the same entity. The perspective is usually fallible. It is answerable to standards of accuracy, well-functioning, and warrant. Since perspectives are ways of perceiving or conceiving, the perspectives are limited by the finite, partial, fallible abilities that they mark or help type-identify.

I take the first thesis as axiomatic here. I believe that it cannot be reasonably denied. I think that it would be absurd to think that finite beings can perceive or think about ordinary objects or properties *neat*. We cannot perceive or think about them without doing so in some representational, perspectival, cognitively limited way. No mental representational ability corresponds to a view that would deny the thesis. We lack cognitive power to perceive or think of ordinary entities in no way at all, or to incorporate them whole into perception or thought—apart from any representational means that constitutes one of many possible perspectives on them, perspectives that mark not just positions in space, but limited perceptual or conceptual abilities.³ Mental representations mark or help type-identify states, capacities, and events. To do so in ways that

3. God was said to have such a power to think of things without any general representation associated with the thinking. The power was called “intellectual intuition.” I regard this view as of doubtful coherence. For present purposes I maintain the more circumspect view that such reference is impossible for finite beings. Their perspective on any entity is limited.

Russell held acquaintance to be the fundamental representational power. He made the mistake of attributing to acquaintance all the key nonperspectival aspects of intellectual intuition except that acquaintance was not in general supposed to bring the objects of thought into being.

Qualitative elements of consciousness are one thing. Singular representation of them (as referents or objects) in thought is another. Treating them as data for perceptual belief is a third. Russell runs these three things together in his notion of sense data. Russell took universals both as properties of objects and as perspectives of the mind on objects. I believe that this is another fundamental conflation. Russell provided no defense of his fantasy about human epistemology and about the mental abilities that go into making reference possible.

All of the foregoing concerns the nature of belief and human epistemology. It seems to me a separate question whether *linguistic* theory can abstract from the perspectival character of thought. Even in this area, I think that the perspectival character of linguistic representation is never fully obliterated in linguistic natural kinds. But this issue will not figure in what follows.

serve psychological explanation, mental representations must type the perspectival, limited abilities that we in fact have.

The main grounds for the thesis derive from reflection on human abilities. There are empirical grounds as well. Psychological explanation takes operations on representations that type mental abilities as fundamental. The transformation and use of representations by perceptual subsystems cannot be separated in empirical theory from the end-product perceptual representations attributed to the whole animal or person, as well as to psychological subsystems.

II

The second thesis concerns conditions on singular, contextual, perceptually based, purported reference. The *main intuitive idea of the second thesis* is that singular, context-bound, perceptually based purported reference must be guided by a general representational content that is attributive. The attributive element marks or type-identifies a representational ability—an ability to categorize referred-to particulars as instances of a type (instances of a kind, property, or relation), and to attribute the type to particulars. The attributive representational content functions fallibly to restrict the perceptually based singular reference to instances of the type. The perceptually based singular reference is to particulars, if to anything.⁴

The second thesis and its companion, the third thesis, are versions of an old idea: Singular reference must be guided by general attributives. My version liberalizes traditional views in two respects. First, the attributions can be perceptual as well as conceptual. Second, the relevant attributed types can be more generic and less sophisticated than the sortal types usually postulated.

Sometimes philosophers sympathetic to Russell suggest that representational contents stand “between” the individual and referents of his thought, and then cast aspersions on such indirectness or mediacy. I think that this is an absurd characterization. Representational contents are ways of thinking or perceiving. There is no alternative to perceiving or thinking in some way, from some perspective. The idea that the representational contents that help type-identify perceptual or propositional states, and that mark those states’ perspectives, are intermediaries, mental objects, screens, or detours between individual and ordinary referent is a product of elementary misunderstanding that rests on cartoon-like philosophizing.

I discuss empirical grounds that support the thesis in my “Disjunctivism and Perceptual Psychology,” *Philosophical Topics* 33 (2005): 1–78. I believe that these grounds overdetermine more general considerations.

4. This main idea extends, I think, to context-bound perceptually based pluralized reference as well. It too must be guided by a general attributive. Context-bound perceptually based plural reference depends on and is grounded in a multiplicity of singular references. I will concentrate mainly on singular reference, discussing plurals only intermittently. I recognize, however, that the thesis has this broader application.

I shall develop these two points of liberalization. The main work of the initial parts of this section is, however, to explain, in much greater detail and precision than traditional accounts do, key concepts that lie behind the main idea. By providing clear and relatively precise explications of the key concepts, I hope to be in a position to argue for the second thesis in an illuminating way. This argument will occupy the penultimate subsection of this section II.

I turn to the explication of some key concepts. This explication will be rather extensive. I ask the reader's patience. A full statement of the second thesis, and the argument for a restricted version of it, will employ the explicated concepts.

Attribution and Singular Application

The key notion in the second thesis is *attributive*. I take this notion as primitive. An *attributive* is a representational content that constitutes a particular way of representing and attributing a kind of individual, a property, or a relation to particulars or to other entities. Any given attributive is one of many possible ways of attributing whatever it attributes. It is a mode of presentation of what it attributes. Attributives are *general* types of representational content. The specific respects in which attributives are general will be the topic of detailed discussion shortly.

Attributives take different forms in thought and perception.

Attributives in thought are predicative concepts. As I use the term "thought," thought is always propositional. So attributives in thought are always components of propositional structures. I assume as evident that every (propositional) thought contains some predicative concept.

Perceptual attributives are general elements in perceptual representational content that type purportedly perceived particulars as being of kinds, or as being or having properties, or as being or entering into relations. Perceptual attributives are what allow perception to be perception *as*, or *as of*. Perceptual attributives are general elements in representational content that help discriminate purportedly perceived particulars by characterizing purported aspects of them. Every perception contains some perceptual attributive or attributives.

One might perceive a particular individual as a body, or as red. Or one might perceive an instance of red as (an instance of) red. Or one might perceive one individual body as being larger than another. Or one might perceive an instance of the relation being next-to as such.

Veridical or accurate perception is always of particulars. *All* perception (perceptual representation)—veridical or not—functions to be of particulars. Since perceptual content, like thought content, constitutes conditions on veridicality, each perceptual representational content must contain one or more singular elements.

The second thesis will claim that all perception functions fallibly to attribute a kind, property, or relation (whether veridically or not) to each

of the particulars that its singular elements purportedly pick out. Perception is always perception *as*, or *as of*. Each singular element in perceptual representational content is guided by a general, attributive element in the representational content. Context-dependent plural representations, which in perceptually based thought are basically groupings of singular representations, are also guided by a general, attributive element.

The second thesis will place a restriction on certain purported representations of *particulars* in certain sorts of thoughts and in all perceptions. What types of *particulars* can be perceived or thought about? Neither thought nor perception is always of individual objects, even when veridical. Particulars include individual objects, events, (particular) masses or stuffs, surfaces, property instances, and relation instances. Ontology does not matter very much for present purposes. Accounting for perception requires, I think, at least these types of particulars. Other types may be relevant as well.

The second thesis will claim that wherever perception purportedly singles out a particular, perception also functions to attribute to the particular a kind, property, or relation. A similar point applies for perceptually based thought. For example, one can perceive or perceptually think about a given instance of the property *red*, correctly, as being (an instance of) red. Or one can perceive or perceptually think about a given instance of the property *red*, mistakenly, as being (an instance of) orange. These purported singlings-out are marked in the representational contents of perception and thought.

In principle, an attributive can fail to indicate any real *type*—any real kind, property, or relation—just as perception or thought can involve referential illusion in such a way that a singular element fails to single out a *particular*. So in perceptually based representation, there can be referential illusions about particulars and attributive illusions about types. In the latter cases, the attributive cannot succeed in attributing a kind, property, or relation. It can still purportedly attribute, or function to attribute, or occur attributively. (I use these phrases interchangeably.)

Thus one can perhaps think of oxygen as phlogiston. Let us suppose, what I think is correct, that there is no such property as phlogiston. There is a phlogiston-attributive that functions to attribute, or occurs attributively. One thinks a thought. The attributive in the thought is predicatively applied. We are supposing that it is predicatively applied to oxygen. But no property or kind is actually indicated, and no property or kind is actually attributed to oxygen.⁵ Still an attributional thought about oxygen has been thought, and the thought has used the attributive *phlogiston*—an attributive that does not indicate or attribute any property.

5. Failures of property or relation indication occur more rarely in perception than in thought. But the details are unimportant here. What is important for our purposes is that the reader maintain a clear distinction between the perceptual attributive (a certain type of representational content), what it indicates and attributes (a kind, property, or relation—an attribute), and what it attributes something *to* (a particular).

The second thesis will be stated for *perceptually based* representation. *Perceptually based representation* comprises perceptual representation and certain representation in propositional thought that is intuitively grounded in perception. More specifically, a *perceptually based representation* is a perceptual representation, or a perceptual memory, or a perceptually guided actional state, or an intermodal non-propositional perceptually grounded state, or an empirical propositional thought (or component of such a thought), or any other psychological state or event *that purports to represent a specific particular through perceptual resources*.

The singular elements in perceptually based representation are certain kinds of *applications*. Such applications are context-bound representational contents that are *individuated in terms of specific occurrences in time*.⁶

A singular application in a *perception* is an occurrent aspect of the perception whose function is to refer to a particular. A singular application in a perceptually based representational content in *thought* is a context-bound element in the thought content that is individuated in terms of an occurrence in time and whose function in the thought is to refer to a particular by way of perception. Such applications are the singular elements in the representational contents of perception or thought, alluded to six paragraphs back. I will say more about singular applications later.

In perceptually based representation there is a phenomenon of *specific plural context-bound application*. Perhaps in perception and certainly in perceptually based thought, there is a phenomenon of representing those *Gs*, with the plural *those* applied to *specific* purportedly perceived particulars. I believe that *in perceptually based representation* such pluralized applications are grounded in multiple specific singular representations. To *perceptually* represent some dots with the representational content *those dots* (where this representation is contextually applied), one must be perceptually representing *each* dot represented by the pluralized representation. The context-bound, perceptually based pluralized representation is *not* equivalent to all the dots, where one allows generalization

6. I will use “application” primarily for singular context-bound applications, and I will often not qualify “application” with “singular” even though I intend singular applications. I take all singular applications to be context-bound.

There are also pluralized demonstrative applications. And I do occasionally use the term “application” for a closely related phenomenon (as I did in the preceding paragraph of text) that is not a type of reference—singular or plural. I use the term for *predicative application*, or more generally *attributive application*. *Attributive application* is an occurrent exercise of attribution (or an attribution individuated in terms of some occurrent exercise)—as distinguished from an occurrent exercise of context-bound singular reference. It is part of the point of the second thesis that all singular application occurs together with attributive application. It should be borne in mind that not all singular applications are perceptual or perceptually based. Some, like applications of *I* or *now*, do not depend for their reference on perception at all. In this section, I concentrate entirely on perceptually based representation.

to be restricted to the dots in a certain region, including perhaps some that are not individually perceived. One must perceive each dot that gets referred to in perceptually based pluralized representations.

Issues over plurals are complex and delicate. I will not attempt to provide a separate discussion of pluralized perceptually based applications. I think that the second thesis could be broadened to include them. That is, each pluralized context-bound perceptually based representation must be guided by a general attributive. But I will focus on singular representation, since I think that in *perceptually based* representation, plural context-bound application is grounded in singular context-bound application (cf. note 6).

The singular elements in perceptually based representation that must be guided by attributives are of two sorts. They include the context-bound singular applications in perceptions or perceptual memory that purport to single out perceived particulars. They also include the applications of demonstrative or indexical elements in thought guided by perception. Singular representations in pure mathematics are laid aside for purposes of this section. They are not, I think, perceptually based.

A more interesting exclusion concerns a type of singular, applicational element that occurs in all perceptually based representation—applications of *de se* markers or egocentric indexes. *De se markers* or *egocentric indexes* are indexical representations that meet two conditions. When applied, they represent an origin for a representational framework, such as a spatial or temporal origin from which the individual's perception occurs. They also mark the origin as of immediate ego-significance for the individual's motivation or for the wider perspective of the individual.⁷

I make applications of *de se* or egocentric indexes exceptions in the second thesis. Although these singular elements do single out particulars, I will not claim that they must be guided by attributives. Their references are held in place by their framework roles in a system of representation.

But singular applications of them on particular contextual occasions are not unrestricted. The applications are not atomistic or "bare." Applications of *de se* or egocentric markers are restricted by their position in the whole framework of coordinates and of attributions that they provide origins for. They could be regarded as a special case of a more general reciprocal dependence of context-bound singular elements on general attributions. But in formulating the second thesis, I want to make explicit the specialness of the case of *de se* markers or egocentric indexes. So I bracket them as far as the second thesis goes.

7. For discussions of egocentric indexes or *de se* elements in perception, see my "Perceptual Entitlement," and "Memory and Persons," *Philosophical Review* 112 (2003): 289–337.

Guidance

The *main intuitive idea* of the second thesis appeals to a notion of *guidance*. I take the notion of guidance as primitive. Still, we can give a rough characterization. An attributive *guides* a context-bound singular representation if, according to the representational content of the individual's overall representational perspective, the attributive is veridical of the particular purportedly referred to by the singular representation; and the attributive is used by the individual or his representational system as an important restriction on the singular representation's purported referent.

Of course, "important restriction" is vague. The relevant important restriction is intuitively an attribution of an *explanatorily significant type*. The attributive categorizes or sorts. The idea is that guidance by way of attribution of types enters into explanations of context-bound acts of reference, both merely purported and successful.

I think that called "Bill", perceived at some time, and grue are examples of representational contents that do not indicate explanatorily significant types that could guide perceptually based singular representation. They could not enter into explanations of context-bound acts of reference. They could not *guide* context-bound singular applications.

There are further restrictions on what sorts of types can guide singular reference through general attribution. Recall that guiding types can be kinds, properties, or relations. There are restrictions on mixing these types in order to yield guidance. For example, I think that attribution of the relation next-to cannot suffice to guide singular reference to instances of bodies or instances of redness. Attribution of relation types can only guide context-bound singular reference to relation instances. Thus a perception of a relation instance can be categorized as being of a relation type next-to or larger-than. But an individual or a property instance cannot be categorized *merely* by attribution of relations that it is perceived as being in. I leave open whether to perceive a relation, one must perceive an entity in the relation.

Similarly, attribution of a property type can guide context-bound singular reference to property instances, but cannot guide such reference to individuals or particular masses. For example, attribution of property types like redness and rough-texturedness cannot alone guide perceptually based reference to a body or a pile of material.

Again, attribution of individual-kind types cannot guide singular reference to property instances, though the veridicality of an attribution of an individual-kind type might entail the instantiation of certain properties constitutively necessary to being an individual of the relevant kind. Attribution and guidance are more fine-grained than entailment.

I do not try to work out a definite notion of categorization or guiding type here. I think of these notions, like the notion guidance, as primitive. The reader should note, however, that the restrictions that I have cited

are quite generic. Empirical reference to individual bodies, for example, does not require guidance by a sortal attributive that applies to middle-sized familiar kinds of bodies and that carries clear count criteria. I will return to this liberality in the account of necessary conditions on attribution and reference in section III.

I have been explicating the notion of guidance in the second thesis's claim that singular, context-bound, perceptually based reference to particulars must be guided by a general attributive. I do not, of course, hold that the representational power and content of such singular reference is exhausted by the guiding, general representations. I have long held that singular, context-bound reference is primitive and ineliminable in terms of conception, description, perceptual attribution, or any other general representational resources. The second thesis itself does not even claim that the nonschematic, semantically general representation must actually apply to, or be true of, the entity represented. Guidance has to do with the functional importance that the representational system must *accord* a general representation in determining the purported referent of the context-bound singular representation.

Attribution and Four Sorts of Generality

Central to the second thesis are certain notions of generality and context-dependence. I believe that the tradition of thinking about attribution has not clearly distinguished these types of generality. Distinguishing them is, I think, critical to understanding psychological representation. In this large subsection, I will explicate these notions of generality and how they bear both on understanding attribution and on understanding how attribution restricts singular, context-bound reference.

The reader mainly focused on the bigger picture, and less interested in detailed understanding, can try to hold in mind the main ideas of the second thesis and move to the argument for a restricted version of the second thesis (two subsections hence) and to section III. I think, however, that a firm foundation for understanding the relation between context-bound singular representation and attribution requires a firm understanding of the relevant types of generality. This section will have some of the character of philosophy of logic. I will try, in some depth, to distinguish notions of generality that are easily conflated.

I will distinguish four sorts of generality in mental representational content. None of these sorts is quantificational generality. One is the generality that concerns the kind of ability (partly) type-identified by the representation. It bears on whether the ability is individuated independently of any particular, specific exercises of it. A second is the generality that concerns how a representation applies to a subject matter—whether by its form and content it can apply to any number of satisfiers or referents. A third is a kind of “syntactical” or logical-functional generality. Finally, there is a kind of generality that requires by its content

a context-dependent act or occurrence in order to apply to a particular. This generality is open to, but unspecific with respect to, context-dependent reference. These are very abstract, overview characterizations of the kinds of generality that I will distinguish in this subsection. I turn now to more specific characterizations.

The first sort of generality, *ability generality*, pertains to types of representational content that mark general, freely repeatable representational abilities. The abilities and the representational contents that mark them are not constitutively dependent for their identities, or for their *relations* to what they represent, on any *particular, specific* set of token applications or representational events (whether these are attributional applications or singular applications). They are not simply abstractions from some particular, specific token application(s) or representational event(s).⁸ They are individuated, and may be learned or innately “wired in,” through acquiring or inheriting a *kind* or *type* of ability. Commonly these abilities are geared to situations or entities of a given type. Such abilities may be—and usually are—dependent for their presence on being constitutively associated with some token applications or other. But if a representational type marks a general ability in this sense, any exercise of an appropriate kind would do. There is no *particular, specific* application or representational event, or any *particular, specific* set of applications or representational events, to which the relevant abilities, and the representational contents that mark them, are *essentially* tied for their individuation, or their relation to what they (purportedly or actually) represent. Individuation goes through a pattern-based type of ability.

All perceptual attributives—representations of kind, property, and relation types—are general in this sense. For example, a perceptual attributive marking an ability to perceive something as a body, as red, or as larger than, is general in this sense. The concepts body, hydrogen, cylindrical, piano, three, tall, malicious, brother of, identical with, next to, the tallest spy ever, and the number 3 are also general in this sense.

I call such representational contents “*ability general*,” since they type general psychological abilities. Such abilities are *freely repeatable*: There are no specific, *particular* token exercises or applications by reference to which the standing representational ability is individuated.

Ability general representational contents contrast with representational contents that mark a *token application* (or *applications*) by some individual perceiver or thinker, purportedly to a particular. I call both relevant acts or occurrent events (and abilities individuated in terms of

8. Which initial event or events count as the attachment of a name to an individual might not matter in socially shared cognition. Who is the first person to start an anaphoric chain of demonstrative reference to some putative particular witch will not matter to individuation, as long as a specific, contextually local set of events grounds subsequent demonstrative applications that go back anaphorically to those specific events.

such acts or events) and relevant representational contents "*applications*." We shall focus entirely on singular applications. Context will make clear which is meant—(a) act, event, ability, or (b) representational content marking an act, event, or ability—if the distinction is important. An application of a demonstrative construction in thought is *not* freely repeatable. The representation, or representational content, marking the act is *not* ability general: There are particular occurrent acts (or events) that are constitutive to the individuation of the occurrence or ability that the representation marks.

A token application of a demonstrative-like construction in perception, language, or thought is to be strictly distinguished from the standing demonstrative construction itself. Thus a representational content marking a token (act) application of the expression "that," or of the standing demonstrative mental representation that, is to be strictly distinguished from the expression "that" and from the standing mental representation that. The standing mental representation that is ability general. The ability to use the demonstrative "that" and the ability marked by its standing counterpart in thought (that) are freely repeatable: No specific event is essential to the individuation of the ability to use and understand the demonstrative construction "that" or the counterpart standing demonstrative mental representation that. By contrast, an application representation is not ability general. It marks a specific act or event. The application act or event itself and any ability, or exercise of an ability, individuated in terms of such an act or event—for example, an anaphoric or memory ability—are not freely repeatable.

Applications purportedly to particulars may be acts in thought—singular applications guided by concepts. Or they may be events in perception—singular applications, purportedly to particulars, of ability general perceptual attributives. Let us call representational contents that mark such acts or occurrences "*ability-particular*" (or "*context-bound*").

It is sometimes plausible to *identify* a singular application representation with a mental act or event. But ability-particular or context-bound representations—representational contents that mark applications—need not themselves *be* token acts or events. They may be abstractions that mark an act or event. Or they may mark an ability or act-type partly individuated in terms of a specific act or event. Although they must be individuated in terms of some particular, specific token application act(s) or event(s), they can be maintained or multiply instantiated over time. A representation that marks the application of a demonstrative in thought can be retained in memory after the token act that helps individuate the representation is past. And if the memory is invoked, the same application occurs again, purportedly to pick out the same particular by way of its anaphoric-memory relation to the original occurrent event of application. An ability-particular (context-bound) representation in thought can be maintained across thinkers, through interlocution. Preservation of context-bound representations, in both memory and

interlocution, has an anaphoric character. All such representations type-identify abilities individuated in terms of particular, specific token acts or events—not in terms of freely repeatable general abilities.

Token singular representations in thought are actively embodied by *particular token applications* of demonstratives like that or indexicals like *I*, and by pronominal back-references taking such applications as antecedents. As indicated, there are analogous context-bound singular representations—individuated in terms of token occurrences, if not acts—in perception.⁹

Paradigmatic concepts are ability general.¹⁰ Attributive perceptual representations are, I think, always ability general.

I now turn to a *second kind of generality*. Most concepts and all perceptual representations that are ability general are general in a further sense. Most concepts and all ability general perceptual representations are capable, according to their form and content, of referring to, being true of, or being accurate of, an indefinite number of entities. Let us call such

9. I discuss this singular sort of context-dependent representation, insofar as it occurs in *thought*, in "Belief *De Re*," *Journal of Philosophy* 74 (1977): 338–362, reprinted in my *Foundations of Mind* (Oxford: Clarendon Press, 2007); "Russell's Problem and Intentional Identity," in *Agent, Language, and the Structure of the World*, ed. James Tomberlin (Indianapolis: Hackett, 1983)—where I introduce the term "application"; and "Vision and Intentional Content," in *John Searle and His Critics*, ed. E. Lepore and R. V. Gulick (Cambridge, Mass.: Basil Blackwell, 1991). The idea is, however, present in my "Reference and Proper Names," *Journal of Philosophy* 70 (1973): 425–439, and "Demonstrative Constructions, Reference, and Truth," *Journal of Philosophy* 71 (1974): 205–223. I developed the role of applications in representational contents in which there is a failure of reference in "Russell's Problem and Intentional Identity." For a focused discussion of singular context-bound applications, see "Postscript to 'Belief *De Re*,'" in *Foundations of Mind*. The first thesis of the present essay is also enunciated in section II of this latter article.

I discuss singular applications as they occur in *perception* in "Perceptual Entitlement." Such perceptual singular elements are needed to account for the fact that individuals perceive particulars, which need not be—and commonly are not—uniquely specified by general perceptual attributions of aspects of the particulars. Individuals' perceptions and perceptual systems represent particular objects and property or relation instances that the perceiver interacts with. They represent those particulars, not look-likes that the perceiver is not interacting with. Analogous singular elements in thought are needed to account for the fact that we can think about objects that we do not fully specify through conceptual representations.

10. I am tempted by the view that all concepts are ability general. One might even take ability generality to be a necessary condition. There are, however, difficult issues here about certain historical proper names. Applying a name like "Aristotle" to the most famous Aristotle requires that one's usage connect to a historical chain that must be characterized in terms of a set of very particular applications. I believe that one's current usage involves an application of a schematic context-sensitive determiner (broadly a demonstrative) that, in use, connects with applications of determiners by other people, ultimately going back to initial applications of the name (or a cognate) to a perceived individual. So the name and the context-sensitive determiner are ability general concepts. But any given application, or file connecting to the chain going back to the most famous Aristotle, is ability-particular. Cf. my

representations “*semantically general*.”¹¹ Most ordinary, noncomplex predicate concepts are semantically general. The concept piano is true of, and open to application to, any number of pianos according to its form and content—even if there were in fact only one piano, or no pianos at all. A visual perceptual attributive representation square is veridical of any visible square entity. A representation is *semantically singular* if its form and content require that it have exactly one referent or satisfier, if it has any. Examples of semantically singular representational contents are the number three, the present king of France (where present is applied), is identical with 3, is the only seven-foot spy ever.

Semantic generality is to be contrasted not only with semantic singularity but also with plural representation that is restricted to specific pluralities of entities. Unapplied, those dots is semantically general. According to its form and content, it can refer to an indefinite variety of dots (in various contexts). But *applied* in context-bound perceptually guided thought, those dots and those 7 dots refer to neither an indefinite number of dots nor to exactly one dot. Similarly, with the concept prime number between 1 and 17. I will count such representations *semantically restricted-plural*.

Syntactic generality is a third sort of generality. Grammar and logic distinguish between singular and general terms. I think that this distinction ultimately rests on representational role. The linguistic distinction has a counterpart in perception and thought: *Syntactically singular* representations are representational contents that function to *refer to* one entity, if to any, when used in a complete sentence, thought, or perception.

“Reference and Proper Names,” *Journal of Philosophy* 70 (1973): 425–439. Whether all noun-like concepts that have this sort of historical specificity can be correctly construed in this way seems to me an open question. Perhaps to learn the concept expressed by “the United States of America,” one has to have been connected to a chain of communication that goes back to a few events associated with the founding fathers. Perhaps not. Whatever the truth of this matter, I continue to think that the distinction between ability general and context-bound representations is an illuminating one. I think that with regard to either most or all concepts, the conceptual abilities are not *explained* or *grounded* in terms of any *specific particular* events. Any events of a certain type would do.

11. I have changed “formally general” in “Descartes and Individualism: Reply to Normore” to “semantically general” here. (The account is also more precise here.) I concluded that the relevant generality is more fundamentally about content than form, though it is associated with both. I do not like the suggestion of language in “semantically general.” I use “semantically” in a broad sense that includes not only relations between signs and what they represent but relations between any representations (including representational contents) and what *they* represent. Here we are concerned with mental representation. In some ways “content general” would be a better term. But it leads to grammatical awkwardness.

In my terminology *veridicality* pertains to both perceptual and propositional representations; truth pertains only to propositional representations. Similarly, veridical of pertains to attributive perceptual and attributive conceptual representations, whereas true of pertains only to attributive conceptual representations.

For present purposes, I take *syntactically general* representations to be those that function (usually fallibly) to be *veridical of* one or more entities. Thus they are representational contents that are predicative or attributive. For present purposes, syntactically general representations are representations whose roles are predicative in thought and attributive in perception or perceptual memory.¹²

A fourth type of generality is what I call "*context-dependent schematic generality*," which for purposes of this essay I will shorten to "*schematic generality*." Representational contents that are or contain indexical or demonstrative representational contents, where the indexical or demonstrative contents are to be strictly distinguished from context-bound applications of them, are *schematically general*, or just *schematic*. To be schematic, a representational content must need a completing context-bound application to have a definite referent or satisfier, and to occur in a perception or in a complete thought. A context-dependent representation unapplied is simply a schema for reference or attribution, applicable in any of various contexts. Even when the context-dependent representational content is applied, it in itself, as distinguished from the representational content that includes the application, is schematic.

Any representation containing demonstrative or indexical elements like this, that, here, now, then, today, I, she, such will lack a definite referent (or, in the last case, satisfier) apart from an application in a context. According to their form and content, they need a context-bound application if they are to occur in a perception or perceptual memory (for perceptual analogs of that, now, or here) or in a complete thought. So any such representation is schematic. Similarly for egocentric

I define semantical generality in terms of satisfiers or referents. A more inclusive definition might cover what some regard as syncategorematic concepts. These are concepts that are ability general, but that do not have a referential or applicational function. Some have thought that quantifiers and logical connectives are examples. They cannot suffice to *guide* context-bound, singular representations.

12. I believe that ultimately syntactic generality is a wider category than I have explained here. Here I take it to be almost equivalent to being attributive. (Almost: I am inclined to think that is such is syntactically general, and predicative, but in itself not attributive. In itself this schematic representational content is not capable of attributing anything.) In perception, the equivalence between syntactic generality and attribution holds, I think, even if one widens the category of syntactic generality. Thus I believe that among perceptual representational contents, all and only syntactically general representational contents—all contents that function to be veridical of—are attributive. In thought, syntactic generality ultimately should be characterized widely enough to include not only predicative concepts but also function concepts (the father of), quantifier concepts, connectives, context-independent determiners, and so on. Not all of these are attributive. I think that the root idea of this wider notion of syntactic generality resides in the notion of there being an open place (for a variable, or schematic representation, or dummy representation) in the syntactic form of the representation that indicates that the syntactical items *operate on* something further. I believe that this is what Frege was getting at with his idea of an unsaturated expression.

or *de se* markers, insofar as they are unapplied or are strictly distinguished from their application.

Schematic representations are commonly complex and contain a mix of elements.¹³ Some elements, the schematic, context-dependent ones, mark a general ability to exercise the completing context-bound application. Some mark a general ability to restrict, in a context-independent way, the context-bound applications. Thus that sofa (unapplied) contains, first, the demonstrative that, which (as unapplied) marks a schematic capacity to exercise context-bound, singular application, and, second, the concept sofa, which marks a general ability to restrict context-bound application in a context-independent way. (The concept marks inferential abilities as well.) The demonstrative that, unapplied, contributes the schematic element to the complex schematic representation. That is purely schematic. There is no restricting element. It is unrestrictedly, completely open to any singular application in any context. The context-independent restricting element in the complex schematic representation (here, sofa) will be called “*nonschematic*.” Both are components of a complex schematic representational content.

These four types of generality will help clarify what sort of attribution must guide singular context-bound perception based representation. But first I want to try to solidify understanding of these different types.

The four types of generality largely cut across one another. They are certainly demonstrably distinct types. To fix and clarify these rather abstract notions, I will give examples that map the relations among the different types. The complex relations among the different types of generality may be more than the reader will want to work through. The argument for the second thesis will rely on a clear understanding of the different types, but the impatient reader may get by with skipping the next seventeen paragraphs.

Ability general representations can be semantically general or semantically singular. The concepts brother, kind, moving, and larger than and perceptual attributives like red and body are ability general and semantically general.

Several types of ability general representations are semantically singular. One type comprises complete definite descriptions (conceptual or linguistic): the natural number that immediately follows one, the human being alive before 2000 with the greatest rest mass, and the shortest spy ever are examples. Another type comprises certain semantically singular, context-independent predicates: identical with 3, natural number

13. There are complications here regarding containment, especially in indexicals, that I shall elaborate later. Note that there are semantically general, syntactically general, schematic representations, like such, next, later. The second thesis applies to applications of these representations. Any application requires a guiding attributive that is nonschematic, semantically general, ability general, and syntactically general.

between 2 and 4, identical with the shortest spy ever, and so on. A third type of ability general, semantically singular representational content, including even noncomplex representational content, comprises individual concepts such as 3, $1 + 2$, God, perhaps Earth, the concept hydrogen, and that 7 plus 5 is 12.¹⁴ Like paradigmatic semantically general concepts, these *individual concepts*—simple and complex—mark freely repeatable psychological abilities. Having the concept is an ability that is not individuated in terms of specific token acts or events of application. It is individuated by a cluster of inferential, applicational, and predicational abilities. These abilities may, of course, be partly individuated in terms of types of application and types of relations to a subject matter. But no particular representational events of application are essential. Yet in each case the representational content must single out one entity if it represents anything. Similar points apply to the other cases of ability generality combined with semantic singularity.

I believe that in *perception* all ability general representations are semantically general. Ability general representational contents that are either semantically singular or semantically restricted-plural occur only in thought. All singular representation in perception is purported context-bound reference to definite particulars. All attributive representation in perception is open to an indefinite number of satisfiers. Restriction to specific, definite satisfiers of attributives in perception occurs only through context-bound representation in response to a cause in context. Such context-bound representation is not ability general.

As far as I can see, noncomplex ability-particular (context-bound) representational contents in both perception and thought are never semantically general. They can be semantically restricted plural or semantically singular, but not semantically general. An example of noncomplex context-bound representational contents that is semantically restricted-plural is an application of a plural demonstrative. The fundamental case of noncomplex, context-bound representational content is singular application. Although they are simple, noncomplex singular context-bound applications never occur apart from a restricting attributive. That is the thrust of the second thesis, which we are leading up to.

Ability-particular or context-bound representational contents can, however, be semantically singular, semantically restricted-plural, or semantically general, if they are complex. The only mother of that person and that body (where that person and that body are applied) are ability-particular and semantically singular. Those dots (applied) is ability-particular and semantically restricted-plural, as is identical with 3 or

14. 3 is to be understood here as grammatically singular rather than as adjectival. I regard 3 as *not* composed of other concepts (such as one and plus). I discuss the point in section V.

Obviously ability general representations can be semantically restricted-plural.

that person (where that person is applied). Person from that city (where that city is contextually applied) is ability-particular but semantically general.¹⁵ I conjecture that all such cases of ability-particularity combined with semantical generality involve complex representations. Applications are combined with semantically general, ability general elements; and the semantically general elements are not within the scope of the ability-particular elements.

Ability generality cuts across syntactic generality. The concept 3 is ability general and syntactically singular. The concept spy and the perceptual attributives square and next to are ability general and syntactically general. That red body (applied either in thought or in perception) is ability-particular and syntactically singular. Next to that red body (a perceptual attributive with that red body applied) and person from that city (with that city applied in thought) are ability-particular and syntactically general.

Ability generality also cuts across schematic generality. Unapplied schematic representational contents like this, today, and that person are both ability general and schematically general. Ability general concepts like ball and ability general perceptual attributives like square are non-schematic. Ability-particular representations like any referential (that is, nonattributinal) application or like that body (including an application in perception or thought) are nonschematic. Ability-particular representations can be schematic if they are complex and only partially applied. For example, that ball is larger than that top, where that ball includes an application and that top is schematic and unapplied, is ability particular because of the application of that ball. But the same representational content is schematic because the context-dependent representation that top is unapplied.

Let us turn to the relations between the second type of generality, semantical generality, and the other types. We have already seen that semantical generality cuts across ability generality.

Semantical generality cuts across syntactic generality. The representational contents body, square, and next to (in perception or thought) are semantically general and syntactically general. Semantically general representations that are syntactically singular are representational contents such as that sofa, the only woman to hit that man, today, he, this, *insofar as they are distinguished from applications of them in a context*. Such representations can apply, by their form and representational content, to various referents. Nothing *in the form or content* guarantees at most one referent. The form and content can be supplemented by contextual application to yield single referents in each context, but indefinitely many referents relative to form and content. It is notable

15. As noted, there are perhaps context-bound semantically restricted-plural representations in perception. There are certainly such representations in thought (*those quartets*, applied to definite quartets).

that all these cases of semantical generality *cum* syntactic singularity are schematic.¹⁶

Examples of semantically singular and syntactically singular representations are that body and now (each including an application in perception or thought) and the concepts 3 and the shortest spy. Semantically singular but syntactically general representations are identical with 3 and uniquely greatest composer ever.¹⁷ Although the matter is complex and speculative, I am doubtful that examples of this sort occur in the representational contents contributed specifically by perceptual systems. Conceptual examples tend to be syntactically complex.¹⁸

What of the relation between semantical generality and schematic generality? Examples of semantically general representations that are schematically general are simple unapplied demonstratives or indexicals like that, now, we. Semantically general representations that are not schematically general include semantically general representations that lack any schematic elements. For example, the attributives body, square, and later than (perceptual or conceptual) are semantically general but nonschematic. Semantically general representational contents that are not schematically general also include semantically general attributives that contain schematically general elements together with their applications. Taller than that man (where the representational content includes application of the schematically general that man) is an example. That man itself is schematically general, but the whole representational content is not.

Obviously semantically singular representations can lack schematic generality—3, the shortest spy ever, and that body (application included). And semantically restricted-plural representations can lack schematic generality—the natural numbers between 2 and 5.

16. There is presumably a *syntactic* category of pluralized definite reference, illustrated by the schematic demonstratives those, they, we, you (plural), and so on. The schematic demonstratives do not have the role either of singular reference or of being veridical-of. Their role is plural reference.

17. I believe that there are no cases of semantically singular, syntactically restricted-plural representations. I take it that syntactically restricted-plural representations are devices for reference not for being veridical-of. But the pluralized syntactical form would prevent reference to a single entity. So either the syntactically restricted-plural is unapplied and semantically general, or it is applied and semantically restricted-plural.

18. Thus in this case, a semantically singular, syntactically singular representation (here, 3) restricts the semantical generality of a noncomplex predicative representation (here, is identical with). In "On What There Is," in *From a Logical Point of View* (New York: Harper and Row, 1953), Quine coined syntactically general terms from singular terms—coined terms like "Pegasizes" or "Socratizes." Quine intended them to be semantically singular and noncomplex. I think that these expressions simply express the complex syntactically general mental representation is identical with Socrates. At least as used in a context (to apply to the most famous Socrates), this representation is an example of a complex syntactically general expression that is semantically singular. It may be important that natural noncomplex syntactically general representations that are semantically singular are hard to come by.

The relations of the third type of generality, syntactic generality, with respect to the second, semantical generality, have already been discussed.

Syntactic generality also cuts across ability generality. Examples of syntactically general, ability general representational contents are simple perceptual attributives, and concepts like person and larger than. Examples of syntactically general, ability-particular representational contents are of such a color and person from that city (each with contextual application included).

Syntactically singular representations that are ability general are complete definite descriptions like the smallest prime and individual concepts like 3 and God. Syntactically singular representations that are context-bound are perceptions like that square body (application included) and applied representational contents in thought like that mathematician.¹⁹

Syntactically general representational contents can be either schematically general (such, such a color, person like her, all unapplied) or nonschematic (body, later than, square). Syntactically singular representational contents can be either schematically general (this, that body, unapplied) or nonschematic (complete definite descriptions, individual concepts, and representational contents like that body—with application included).²⁰

The relations between the fourth type of generality, schematic generality, and the other types have already been laid out.

In developing the second thesis, I am primarily interested in the attributive elements in representational contents scope-governed by *perceptually based singular applications*. As long as all schematic elements are applied (singularly or in a plural-restricted way), the complex representations governed by perceptually based singular applications are nongeneral in all four senses of generality. Since the relevant representational contents are scope-governed by applications, and applications are individuated in terms of particular acts or occurrences, the representational contents cannot be ability general. Since the function of the applications is to refer to definite particulars, they cannot be syntactically general or semantically general.²¹ Since the representational contents include applications of all schematic elements, the representational contents cannot be

19. Analogous cases can easily be produced of syntactically restricted-plural representations that are either ability general or context-bound.

20. Obviously syntactically restricted plural representations can be either schematic or nonschematic.

21. Applications of such are delicate. Such commonly simply functions as an anaphoric pronoun, going back to some predicative concept that occurs earlier. But this syntactically predicative schematic concept (such) can be combined with a demonstrative application. Such combinations are complex representations. I believe that, at least in perceptually guided cases, the demonstrative application is singular or restrictedly-plural. The application refers to an instance, or instances, of a type. The application is guided by an attributive (beyond such) that attributes a type, and such is understood to indicate that type. In

schematically general. Schematic generality is openness to application in various contexts, and hence in itself does not include application.

Applications of *de se* markers or egocentric indexes are special cases of these points. Like other singular applications, they are nongeneral in all four senses. Although such markers do not apply to perceived entities, their applications do establish origins for frameworks of empirical representation. The acts or events of application are themselves particulars, or are individuated in terms of particulars. Hence the acts of applications are not freely repeatable. Since representations that mark acts of applications are individuated in terms of such acts, they cannot be not ability general. An origin of an empirical framework must be a particular. Given that the function of representations that mark acts of application is to represent particulars in a singular way, they cannot be syntactically general. Representations marking events of empirical application cannot be semantically general because their reference, if any, is to particulars that occur in the specific context of the application. Although the egocentric indexes involve schematic elements, the application of such elements in particular contexts (hence the representational content that includes both schematic element and its application) cannot schematically range over any of various entities, and hence cannot be schematically general.

Egocentric indexes are special cases that are not subject to the second thesis, but I want to keep them in view, since their singular reference is restricted in ways that are at least broadly analogous to the ways attributives restrict applications guided by perception.

I am interested in attributive guidance in singular, context-bound, perceptually based *reference*. What do the four types of generality have to do with *attribution*? I believe that the notions of generality can be used to sharpen understanding of the sort of attribution necessary in guiding singular, context-bound perceptually based reference.

All attributives are syntactically general. Their form reflects their function. Their function is partly to be veridical of (perceptually accurate of, or propositionally true of) entities to which they attribute kinds, properties, or relations. Not all attributives are semantically general or ability general. Is a unique smallest prime is attributive but semantically singular. Killed that person (application included) is attributive but not ability general. Noncomplex attributive psychological representational contents *may*, however, always be ability general.

Be that as it may, the specific sort of attributive that I believe is necessary to guide singular, context-bound, perceptually based reference is nonschematic, ability general, and semantically general. That is the

linguistic expressions of perceptually guided thought, commonly the expression “such” is used to evoke a perceptual attributive or a concept containing or grounded in a perceptual attributive. Since such attributives are commonly not precisely expressed in language, “such” is used to evoke them rather than express them.

main thrust of the second thesis. I will argue for this claim later in this section. For now, I will make a few intuitive remarks about it.

The guiding attributives are nonschematic: They mark a general representational ability to place restrictions on applications in a way that is independent of the context of the applications. They do not by their form require a completing application to have a definite referent or satisfier in a perception or complete thought. The satisfiers of the guiding attributive are fixed by the representational content of the attributive. That is how the attributive restricts the context-bound singular reference. More centrally, to purport to represent a particular in a context-bound perceptually based way, one must represent the particular as an instance of a type. Schematic representations in themselves do not do this.

The relevant guiding attributive can, of course, be *contained* in a complex schematic representation. Its contribution to guidance of context-bound singular reference is a purported context-independent restriction of the type of entity that can be a referent. Recall that I in effect defined a *not-purely-schematic* representation as a schematic representation that *contains* a nonschematic representation. Syntactic containment is clear in the case of explicitly complex representations (that sofa). But there are more implicit cases. Most demonstrative pronouns and most indexicals seem to contain some schematic/nonschematic mixture. Thus she is restricted to females. Now and there are restricted to times and places, respectively. We is restricted to persons, or at least beings with psychologies. Today is restricted to days. I believe that these restrictions derive from nonschematic representational contents that are contained within the schematic ones. Thus I believe that applications of today in particular contexts can be *guided* by a semantically general, ability general, nonschematic attributive day.²²

There seems to be some analog between such restricting elements in ordinary indexicals and restricting elements (like ego or place) in *de se* or egocentric markers. But there also seems to be a significant difference between ordinary indexicals in thought and language, even framework-marking conceptual indexicals like here or now, and *de se* markers in perception and thought. Guidance by the nonschematic restricting concepts contained in the conceptual indexicals and ordinary demonstratives in thought seems to have a significantly different role from that of the nonschematic restricting elements in *de se* markers. Guidance by nonschematic perceptual elements that restrict the objects of perception also seems different from the restriction contained in *de se* markers in perception. It is not that there is less restriction. The difference lies in the role or type of the restriction. Although a semantical restriction goes on in *de se*

22. It seems to me important that day and time guide in today and now, whereas near or salient does not attributively guide applications of this (differentiated from that).

markers or egocentric indexes, I doubt that it should be assimilated to attributive *guidance*.

The restriction in the egocentric indexes channels through the role in the framework rather than through helping the application discriminate an instance of a type. One does not perceive or think about the entities (the self or the place) that are marked by *de se* markers or egocentric indexes. Neither perception nor thought “picks out” the parameters indicated in applications of these markers. There is no scope for illusion or error. Yet the semantics of representational content cannot dispense with such markers. Application of such markers is fundamental in establishing the origin for a thinker or perceiver for a representational framework. Such context-bound applications are not representational acts or representational events (e.g., perceptual events) that occur within the framework. Such occurrences are part of the contextual establishment of the framework itself. Even with *de se* markers or egocentric indexes in perception, there are, however, semantically general, ability general, nonschematic restricters. It is just that the role of the restriction seems sufficiently different that I would like to highlight such markers as special cases in the second thesis. I leave fuller discussion of them to another occasion.

The guiding attributives are not only nonschematic. They are also ability general and semantically general. The requirement of ability generality derives from the idea that the singular representation can purport to refer only if it flows from a general ability to represent entities of the type.²³ The requirement of semantical generality derives from the idea that purported context-bound reference to a particular must be purported reference to an instance of a type.

The notions nonschematic, ability general, and semantically general restrict what it is to attribute a type in a way that can guide a singular, context-bound, perceptually based representation—a perceptually guided application.²⁴

Here we have a very substantial approximation to the full second thesis: Laying aside egocentric indexing elements, a perceptually based context-bound semantically singular representational content must, in order even to purport to refer, be guided by a nonschematic, ability

23. The notions nonschematic and ability general help articulate an intuition that representation in perception and in perceptual thought is, in a minimal way, *objective*. The relevant objectivity concerns independence of perspective from a particular context and from particular token acts or events. It is a kind of objectivity that is associated with (relative) nonparochiality of perspective.

24. Being nonschematic, semantically general, and ability general does not suffice for being attributive. Complete general plural constructions like all the dogs ever meet these conditions, but are not attributive. It is at best unilluminating to claim that being semantically general is necessary for being an attributive representation. If is identical with 3 is not attributive, much further explanation is needed. Being nonschematic is also not necessary

general, semantically general attributive. The attributive must be non-schematic to be sufficiently substantive to restrict the reference. It must be ability general if the referential application is to be the exercise of an ability to represent instances of types. It must be semantically general in order to purport to represent a particular as an instance of a type. These claims are very rough. They are meant to be intuitive, not probative. Argument will follow.

Perception and Thought: Association

At the beginning of this section I indicated that the *main idea* of the second thesis is that singular context-bound perceptually based purported reference must be guided by a general attributive representational content. I have been sharpening the relevant notions of guidance and attribution.

The full second thesis is more complex than its main idea. Much of the complexity concerns relations between perception and thought. I think that the ways that context-bound singular representation can be guided by general representation are more varied for thought than for perception. Many of the details of formulation are more conjectural than the main idea. If one wanted to avoid the details, one could skim or skip this subsection, and return to the argument that I give for a restricted version of the second thesis—a version that specifically concerns perception.²⁵

The complication regarding the relation between perception and thought hinges on different ways that a guiding attributive can be *associated* with the context-bound singular representation. Discussion of association is in effect discussion of various forms of guidance. What is packed into supposing that each context-bound singular representation must be *associated with* some guiding nonschematic, semantically general, ability general representation?

A simple case of association is a demonstrative in thought accompanied by a perceptual concept or some sortal. A semantically general representation *accompanies* a context-bound singular representation if (a) the singular and general representations are contained in a single complex representation and (b) according to the representational content of the complex representation, the referent of the singular representation is a satisfier of the general representation.

Containment is a logical-grammatical notion. Psychological states are partly type-identified in terms of their representational content. Their representational content has a certain logical or grammatical organization determined by the ways abilities type-identified by the different elements

to being an attributive representation. Person from her city (unapplied) is schematic but attributive. The notions ability generality, nonschematic, and semantically general are intended to sharpen the *sort* of attribution involved in guiding singular reference. I take the notion attributive to be fundamental and primitive.

25. The term “accompanies,” about to be explained, does occur in the argument.

of representational content interconnect psychologically. In an occurrence of the thought that pigeon sees the kernel, the singular element that consists in the context-bound application of that is accompanied by the semantically general representation pigeon. Pigeon is contained in the representational content that pigeon (and, more generally, in the thought), inasmuch as the logical form of the thought helps mark how essentially separable psychological abilities (the one associated with the schematic that and the one associated with pigeon) are related in the complex representational content of a thought. In this case, the logical-grammatical relation of containment also marks a relation between the semantical roles of the two components. The complex that pigeon is fully successful semantically only if in application of it, there is a unique referent of that that is a pigeon. Accompaniment is a relatively strong form of association.

Accompaniment and association are relations not only in thought but also in perception. I believe that every *perceptual representation* must either be, or be accompanied by, or contain, some nonschematic, semantically general, ability general perceptual attributive. (I believe that all noncomplex perceptual attributives are nonschematic, semantically general, and ability general.) Perceptual representations commonly contain a multitude of such general representations. In a visual perception as of a body, one part represents the body's left side; another part represents the center of the body's front surface; another aspect of the percept represents the color at a certain place on the body's surface.

All these semantically general representations are *contained* in a larger topological perceptual representation. The larger perceptual representation also contains context-bound singular elements that are *accompanied by* contained semantically general elements. What this (partly) means is that psychologically speaking, the exercise of the singular perceptual application (the application of that) and the exercise of a general perceptual ability (the exercise of the ability marked by side) occur together and are mutually dependent. The "grammatical" accompaniment also marks ways that a psychological system transforms perceptual content either in the formation of new perceptions in response to stimuli or in perceptual memory or perceptual expectation. I think that every context-bound singular element in a perceptual representation—other than framework-indexing elements—must be *accompanied by* some nonschematic, semantically general, ability general perceptual attributive(s).

Constraints on associations between singular, application *thought components* and general representations are looser than those on perceptual representations. Not every context-bound singular element of a thought must be *accompanied by* a semantically general *concept*.

An application of a demonstrative can be associated with general representations by being tied pronominally to concepts in other thoughts by the person. Application of a demonstrative in one thought may be thus guided by concepts in other thoughts. One could think thoughts

that have file-like or anaphoric memory connections back to singular elements in other thoughts. The guiding representations may be present only in the other thoughts. The thinker must, in these cases, be capable of making a relatively immediate connection between the application of the singular representation and an application of the guiding concept.

A singular demonstrative-like element in thought could perhaps be associated with, and be guided only by, a general perceptual or perceptual-memory attributive rather than a conceptual predicative attributive. An individual could perhaps think thoughts containing singular, referential elements guided by perceptual types, but no specific concept. Suppose that one had an unconscious perception of an object but had not conceptualized it. One could perhaps essay a demonstrative reference in thought guided by the perceptual type. The individual might not have conceptualized *in thought* what sort of object it is, how it is sensed, or even in what direction it is. Then no such representational material is available to the individual's *propositional* inferences. The individual might still be able to say, when presented with candidate objects, "that's not it," "that's it." In so doing, the individual might be going on some unconscious perceptual way of tracking it.²⁶

To say that an autonomously applied, context-bound singular representation must be *associated with* a general representation is to say that one of the following conditions holds: (1) The singular representation is perceptual and is accompanied by a perceptual attributive. Or (2) the singular representation is a component of a representational thought content; and (a) it is accompanied by a general conceptual attributive; or (b) it is anaphorically connected in memory or reasoning to singular representations in other thoughts by the same thinker that are so accompanied (as in (a)); or (c) it is anaphorically tied to and guided by a perceptual attributive in the same thinker. In all these cases, the second thesis holds that the singular representation must be associated with a nonschematic, ability general, semantically general attributive, somewhere in the thinker's point of view, that *guides*. The association with a guiding representation is necessary even if (as in cases (2b) and (2c)) the associated general representation does not accompany the singular representation.

Recall that the main idea of the second thesis is that purported, singular, context-bound, perceptually based reference must be guided by attributive, general representations which type-identify types of

26. If the individual just has a feeling that an object is present but *lacks* any perceptual representation, however unconscious, of any aspect of it, then he cannot think a demonstrative thought about a definite object in this context. The example that I have been discussing came from conversation with Calvin Normore. I place no great weight on the idea that an individual might think demonstrative thoughts guided only by perceptual representations. I sketch it mainly to give a sense of a space of putative possibilities.

particulars referred to. The first thesis holds that all representation is perspectival, representation-as.

The second thesis presupposes the first, but makes a more detailed claim. Fully stated:

*Laying aside egocentric indexing elements, each context-bound (ability-particular), semantically singular, syntactically singular, perceptually based representational element in every autonomous propositional thought, and in every perception, perceptual memory, perceptually guided actional state, and perceptually grounded intermodal state must be associated with a nonschematic attributive that is ability general and semantically general, and that guides the singular representation.*²⁷

“Autonomous” is meant to rule out thought that leans essentially on communication for its reference. I shall discuss this qualification in section III.

I want now to give an argument that explains why singular elements in *perception* must be accompanied by and guided by general elements—nonschematic, ability general, semantically general, attributive representations. Perceptually based singular representations in thought and perceptual memory depend on such elements in perception and carry analogous requirements.

Argument for a Restricted Version of the Second Thesis

In this subsection I propose an argument for a restricted version of the second thesis. The argument is restricted to the thesis as applied to *perceptual* context-bound, semantically singular representation (excluding *de se* framework markers). Perceptual representation is trivially perceptually based. It occurs in a perception. The argument derives from considering fundamental aspects of perception and perceptual explanation. I think that the argument can be expanded to apply to perceptually based thought. I shall not do that here.

27. As noted earlier, I think representational contents like God and three are individual concepts, the conceptual counterparts of individual constants. They are ability general but semantically singular. Unlike ordinary proper names, they are not associated with demonstrative-like determiners. Cf. “Reference and Proper Names.” I believe that the second thesis can be modified and extended to apply to individual concepts. I think that these concepts also cannot be thought autonomously unless they are associated with and guided by semantically general, predicational concepts. Thus in autonomously thinking the concept three, one must be disposed to think such things as that three is a number, or three is the number immediately following two. Or one must associate the individual concept with attributive uses accompanying and guided by sortals: there are three Beethoven string trios. Autonomous use of the concept God or Tlaloc presupposes general attributions, such as deity, agency, god of rain. I believe that individual concepts require some nonschematic, semantically general, attributive conceptual associations to enable them to be context-free. Expanding the argument (which follows) for the second thesis to cover such cases would, however, be nontrivial.

Restricted 2nd Thesis. Any perceptual, context-bound, semantically singular representation that is not an application in an egocentric indexing element must be accompanied by and guided by a nonschematic, attributive, ability general, semantically general representation.

I begin, in the two paragraphs that follow, with some background observations about perceptual ability. Then for an arbitrary perceptual, context-bound, semantically singular perceptual representation *S* that is not an egocentric marker, I argue that there must be a further representation *G* that has each of the features that the Restricted 2nd Thesis requires.

To be, or to be part of, an exercise of a perceptual *ability*, an occurrence of a context-bound singular element *S* in a perceptual content must be part of an instantiation of a *general pattern* of perceptual response. The response is triggered by proximal stimulation. It is fundamental to perception, and psychological explanation of perception, that given the perceptual abilities and antecedent psychological setting of the psychological system (type-identified partly in terms of general patterns of relations to usually distal, environmental stimuli), any given type of proximal stimulation would produce the same general pattern of perceptual response.²⁸ In cases of successful perception, the response discriminates entities that are environmental causes. The general pattern of response in which singular element *S* is embedded must be repeatable if it is to be part of an exercise of a perceptual ability, and if it is to be a perceptual response that is the basis for explanation.

To mark a *perceptual* ability, and to be explainable as a *perceptual* response, the general pattern must be type-identified so as to allow for evaluations of veridicality or referential success. It must be type-identified in terms of *representation* (representational content). Call this representation or complex of representations that type-identifies a *general* pattern of perceptual response "*G*."

Individuals perceive particulars through their perceptual systems by representing natural aspects of those particulars. There is no other way. One cannot perceive Mama except by representing some of her natural characteristics. The general pattern of perceptual response functions fallibly to discriminate particulars by way of natural aspects of those particulars. The natural aspects can be represented similarly by any appropriately equipped perceiver on a different occasion. So the relevant general pattern is *freely repeatable*: There is no particular *occurrent* stimulation, proximal or distal, that is essential to its nature. Since it is freely repeatable, the representational content that marks it is *ability general*.

A second line of reasoning leads to the same conclusion. The general pattern instantiated in any given perceptual response is explainable as a type of response produced by a type (or range of types) of stimulus, given

28. This is an approximate formulation of the Proximity Principle. Cf. "Disjunctivism and Perceptual Psychology."

antecedent and cooperating psychological settings. In successful cases, the response is explainable as deriving from a type of environmental *representatum*. It must be so explainable if perception is to be subject to general explanatory principles. The representational identity of the perceptual response-type depends on the types of stimuli that produce it. So the ability that the response-type constitutes is freely repeatable. So G is *ability general*.

The ability generality of G is thus both grounded in the nature of a perceptual ability and signaled by the methods of explanation in perceptual psychology. I have given two arguments for the ability generality of G. The first goes from the (fallible) function of perception in representing particulars by representing natural properties to the nature of the perceptual abilities (*viz.*, freely repeatable) and the nature of the representations that mark those abilities (*viz.*, ability general). The second argument centers on the nature of the explanation of perception. The second argument is from a weaker premise. It simply notes that explanations of perception take (I think *must* take) as a fundamental assumption that the perceptual response, whatever its function, is a response to stimuli of certain types. There are no specific *particular* occurrent stimuli or responses that the response is individuated in terms of. This is part of what it is to be a natural response.²⁹

Since the pattern of response is freely repeatable and not essentially or necessarily causally linked to any particular token stimulation—proximal or distal—but only to any stimulation of a range of types, and since the reference and veridicality of perceptual representations that type-individuate perceptual ability depend on environmental causal relations with the subject matter, G is *semantically general*. If G can represent or be veridical of a given particular, it could represent or be veridical of any other particular of the same type, as far as G's form and content are concerned. Such form and content apply to any of various instances of properties or relations, or to any of various particulars of a kind.

The relevant type of successful application of G is veridicality-of, not reference. The function of perception is to discriminate particulars that the perceiver interacts with. Discrimination must be by way of general perceptual abilities. These abilities are repeatable, but they are of perceptual relevance only insofar as they enable the perceiver to discriminate particulars. So the abilities respond to repeatable patterns in helping to discriminate particulars. The form of the representational contents that help type-identify these abilities is determined by the abilities' function. So the form of the semantically general representation G (or component of G) that marks the system's capacity to discriminate context-independent aspects of the environment is *attributive* rather than referential. G's

29. The stronger premise of the first argument will be reused in the argument that G is nonschematic. The premise of the second argument is not strong enough to show that G is nonschematic. It suffices, however, to show that G is ability general.

role is to help perception—representations like S in perception—represent particulars as being of certain types or as having certain repeatable aspects.³⁰

Again, I argue from function to content. Perception cannot represent general properties “in the abstract.” Semantically general representations in perception can occur only in tandem with singular representations. Perception functions (fallibly) to represent particulars via those aspects of them that it can discriminate in the context. This pattern of interdependence is marked in the “syntactical” or “grammatical” organization of perceptual representation, which treats singular and general representations as contained in a unit representation. The complex representation fulfills its function if the referent of the singular representation satisfies the general representation. So the general representation or representations G *accompanies or accompany* the singular representation S. (We could take the representations to be a subset of G. But I think that the argument is not compromised if we identify them with G.)

Since perceptual explanation appeals to G along with S in explaining instances of perception, G must be relevant to explaining the perceptual system’s purported discrimination of what S purportedly refers to by way of purportedly discriminating aspects of that referent. A similar point derives not from the nature of explanation but from the nature of perception. The nature of perception requires that G type-identify an ability to respond to such discriminable aspects of the environment. Purely schematic representations do not mark a response to any specific type of stimulation or to any specific aspects of the environment. They cannot mark what is specific to the perceptual system’s discriminative response. If S were backed purely by a demonstrative ability marked only by this, the ability would not be a specific discriminative response to any of the numerous proximal and distal particulars, or types of particulars, in any given causal chain from the environment to the sensory receptors. Such a *this* would not mark an ability that was specific to the perceived environmental particular, even in the context. For the *this* would mark an ability that was equally (un)specific to the array of proximal stimulation, the perceived entity, the entity’s surface, any of its various properties, the stimulus patterns at any of various distances between the perceived entity and the perceptual system, and so on. G must mark an ability that is both

30. I take it that the notion of attribution derives from Aristotle’s notion *saying of*. Here we are not discussing saying. We are discussing any sort of mental state, event, or act, with the form and function of connecting particulars to general types (attributively!). I believe that the same idea, confined to a perception-like case, is present in Kant’s notion of a predicate of intuition, *Critique of Pure Reason*, B278. In my terminology, predicates are *conceptual* attributives, but I think the difference from Kant is only terminological. He recognizes a distinction between perceptual attributives (“predicates of intuition”) and conceptual attributives (“predicates of judgment”).

specific and repeatable. It must specify some purported aspect of the perceived entity. So G is not purely schematic. Only restrictive, nonschematic aspects of impure schematic representations are relevant to marking the perceptual system's capacity to discriminate particulars by way of their aspects. Only they mark capacities to respond to and discriminate recurring, non-context-specific aspects of the environment. So either G or a component of G must be *nonschematic*.

Thus any demonstrative in perception must be accompanied, in purporting to secure its referent, by a nonschematic qualifier. This (as applied), or even this is G (as applied) is not a possible form for perceptual representational content. Only this G (as applied) will do as a purportedly referring perceptual content.

Since individuals' perception and their perceptual systems function to pick out particulars through singular representations and accompanying general representations, and since the general representation G has no other perceptual function than to apply to particulars so picked out, G, or a component of G that meets the conditions discussed above, is for the perceptual system an important means of helping to pick out the particular. Given our *official* explication of guidance, it follows that G (or a relevant component of G) *guides* the singular representation S.

But our supplementary remarks about guidance went further. I stated that guidance was by way of some sort of explanatorily significant typing. I maintained that certain types of not-purely-schematic attribution are insufficient. On the other hand, I left open exactly what would count as an explanatorily significant typing of a particular. Can more be said here?

Let us consider the matter intuitively. Insofar as perception is, purportedly or actually, of an instance of a property or relation (a color or a spatial relation, for example), it seems necessary in producing the relevant typing of instances that some property in the same range be attributed. So one could (purportedly or actually) perceive an instance of red *purely* as red or as orange, but not *purely* as flat, or as an edge, or as a smell. What funds this intuition?

A perceptual system, or a perceiver, must have abilities to discriminate particulars by way of types that the particular instantiates. The system, or perceiver, must be able to discriminate those types. I believe that *the key underlying principle that informs the search for guiding types* is as follows: The ability to discriminate a particular must be marked by some ability general attributive representation that under certain normal, standard-making, conditions would be successful in helping the perceptual system discriminate the perceived particular from discernible particulars of other types that are in the same environment. Moreover, this success must ground explanation of the nature of the perceptual state and its representational content through explanation of the relevance of perception to the individual's basic activities. (I take this last condition to be assumed in what follows.)

Of course, since shapes are of different colors and colors commonly color a wide range of shapes, neither attributive type would be a good general

guide for finding instances of the other. But even if shapes and colors correlated more closely than they do, an attribution of a shape alone could not be a means of perceptually discriminating a color—because shapes and colors are both in the environment. So an ability general attribution of shape could not by itself discriminate an instance of a color from an instance of a shape. Attribution of shape would not in normal conditions of successful perception suffice to discriminate an instance of a color from various shapes that are equally in the environment.

It is perhaps less evident what to say about perceptual kind-typing of individuals, such as bodies and events. Following our conjectured principle: The perceptual system must attribute kinds that under the normal content-determining conditions would be successful in discriminating the perceived particular from particulars (including concrete individuals) of other kinds in the same environment. Perceptual representation as of bodies must discriminate them, in normal conditions, from events, color instances, and shape instances—because these kinds of particulars are distinct and occur in the same environment.

Of course, one cannot perceptually discriminate bodies apart from *any* aspects of them. There is no perceptually discriminable kind body that is perceptually identifiable independently of some aspects of bodies. Traditionally, certain sorts of trackable shape and perhaps solidity or resistance to touch have been regarded as the key aspects. This is a complex matter that I will not try to unravel here. I want to make a few more remarks, however.

I think that some generic type of volume that is discriminable from the background and stable enough to track is the key feature that triggers perceptual categorization of something as a body. But any instance of the generic shape or volume type—such as being coherent, relatively rigid, three-dimensional, and largely bounded—that is a mark of perceptible bodies must somehow be distinguished from the object itself. Inevitably, for any generic three-dimensional shape that correlates with bodies, the shape and body cannot be perceptually discriminated in a given case.

I think that the distinction lies in the function that the body representation has in the representational and practical economy of the perceiver. A perceptual representation of a particular as a body, as opposed to a generic shape that is the perceptual mark of a body, has certain representational functions in unifying or binding representations of various sorts—shape representations, color representations. Moreover, it is a necessary aspect of the ability to discriminate bodies (and only a contingent aspect of the ability to discriminate the relevant shapes that one uses to discriminate bodies) that the perceived particular be trackable over time. A body representation is also essentially connected to practical functions such as eating, mating, fleeing, and avoiding collisions, in ways that the shape representation is not.

The association of these functions with perception can ground explanations that take both guiding body representations and whatever (also

commonly, guiding) generic shape representations on which guiding body representations are necessarily parasitic, to coexist in a perceptual system, and yet to be distinct. It is necessary to the body representation but not the shape representation that its application be associated with an ability to track the particular instance (of body), bind it with other property-representations, and connect it immediately with certain practical functions.³¹ But the body representation is representationally (though not necessarily temporarily) posterior to the generic shape representation. In the order of explanation of the formation of representations in a perceptual system, the shape representation must come first. I reject the view that conceptual representations in propositional attitudes must supplement perceptual representations for the distinction to have empirical application. This is a complex matter, not to be argued here.

This line of reflection suggests that guidance is in terms of typing attributions that provide an explanatory ground for distinguishing successful perception of particulars of a given kind from perception of particulars of other kinds that are also present in the environment. The ground may lie only in perceptual discrimination. Or it may also involve further functional differences in the abilities marked by the representations. A full account of typing attribution must attend to empirical detail.

I have argued that any perceptual, context-bound, semantically singular representation that is not an egocentric indexing element must be accompanied by and guided by a nonschematic, attributive, ability general, semantically general representation.³² The argument is meant not only to support but to clarify this conclusion.

Two Brief Comments and Three Groups of Remarks on the Second Thesis

I will make two brief, retrospective comments on the second thesis, followed by three more extensive observations about it.

First, I do not count the widest category, *object* or *entity*, as sufficient to guide singular elements in perception. The perceptual system could not have a representation that purports to apply to numbers or thought events as well as physical entities. The point of the general element in perceptual representation is to characterize what it is about the purportedly perceived particular that makes it perceptually discriminable from other particulars in the environment. The most generic categories lack such a function.

31. Cf. Anne Treisman, "Feature Binding, Attention, and Object Perception," in *Attention, Space, and Action*, ed. G. W. Humphreys, J. Duncan, and A. Treisman (Oxford: Oxford University Press, 1999).

32. The requirement of accompaniment holds only for perceptually based representation. The second thesis relaxes this requirement for thought in general.

Second, I do not require that the general representation be conscious. Many of the primitive representational aspects of perception are preconscious.

Now to the *first of the three more general groups of remarks*. The second thesis is incompatible with holding that a perceptual context-bound singular representation can fail to be guided by a nonschematic general representation. For example, a sensation *per se* or a neural disturbance *per se* does not suffice to recruit a singular perceptual demonstrative.³³

It is, of course, an empirical matter whether any given system is a perceptual system. What rules out sensations and neural disturbances as kinds fully adequate to the task of marking psychological aspects of singular *perceptual* representation of entities in the physical environment is their lack of referential specificity. The point emerges from considering what psychological representations must be in place in the perceptual system's response if perceptual discriminative abilities that are marked by

33. In some of his articles on multiple object tracking, Zenon Pylyshyn skates very close to the view that I am arguing is incoherent. Often Pylyshyn seems only to be claiming that the reference of the demonstrative indexes is not determined by coding properties of objects that distinguish them from other objects in the scene. (All of the dot-like entities look the same.) But he sometimes suggests that visual indexes are recruited without being associated with an encoding of *any* properties of the objects that they track. Sometimes visual indexes are said to be preconceptual and to pick out "visual objects" that are "proto-objects." It is unclear to me whether Pylyshyn intends "preconceptual" to amount to "preperceptual" (in my sense). Some of what he writes suggests that he does. On this view, the visual indexes start as no more than sensory registrations. Sometimes it is suggested that the visual indexes do not genuinely refer. They would thus not count as "objective" or as perceptual in my sense, and the theory would not be contradicted by the second or third thesis. (For a general sketch of my conception of the *perceptual*, see my "Perception," *International Journal of Psychoanalysis* 84 (2003): 157–167.) I find Pylyshyn's discussions of visual objects, like most such discussions in the vision literature, unclear. By the stage at which visual indexes are involved in *perceptual* tracking, I believe that they must be associated with representation of at least primitive properties. Some of Pylyshyn's actual discussion of the indexes in experiments treats them as picking out objective entities. The moving objects on a computer screen clearly *model* physical objects. I believe that the objects on the screen are fully objective, and that there are, at this point in perceptual processing, genuine perceptual, context-bound singular representations. Construed in this way, Pylyshyn's apparent claims that no properties are encoded when visual indexes are used seem to me clearly mistaken. Indexes in genuinely *perceptual* tracking initially pick out objects not only by way of spatio-temporal position but also in terms of some sort of minimal, approximate boundedness or integrity of the form of the objects—some way of distinguishing figure from ground. They are visually discriminated from their backgrounds by some generic, approximately closed geometrical forms, though under time pressure not by their specific shapes. This is true, I believe, even if something like a visual index is initially recruited by a reflexive nonperceptual registration of simply a sudden onset of proximal stimulation. Although shape, color, and kind are sometimes not tracked when objects are tracked, if the indexes pick out genuinely *perceived* objects, they must (and do) carry minimum coding of a perceivable type, however generic, that distinguishes figure from ground. Cf. Zenon Pylyshyn, "Connecting Vision with the World: Tracking the Missing Link," *The Foundations of Cognitive Science*,

those representations are to be explained.³⁴ Such explanation must explain responses to stimulations by attributing perceptual competencies to get things right. Full explanations purely in sensational or neural terms do not appear viable. Such terms are representationally too unspecific. Sensations and neural disturbances cannot in themselves supplement applications of singular perceptual demonstratives to explain perceptual ability, perceptual reference and veridicality, or perceptual error.³⁵

The lack of referential specificity of purely schematic or purely singular representations forces guidance by nonschematic, semantically general representations. The requirement underlies the fact that perception—both singular and general perceptual representation—has conditions for success in objective reference and veridicality.

The point of explanations in perceptual psychology is to explain how the perceptual system represents particulars and aspects of the environment—and ultimately how it enables the individual to perceive what he, she, or it does perceive—under various types of stimulation.

Any perceived particular is of given types and has given properties. Other particulars with other properties are present in the causal chain that triggers the singular representation. In fact, for any pattern of causal relation between a perceived particular and a perceptual response that succeeds in representing the particular, there are many kinds of particulars that are necessary parts of this pattern but are not perceptually represented. For vision, there are events on the object's surface, light waves in the space between the perceptual system and the particular, chemical changes on the eyeball and retina, events in the optic nerve. Each of these events is part of a pattern of events, any instance of which

J. Branquinho ed. (Oxford: Clarendon Press, 2001)); Pylyshyn, "Situating Vision in the World," *Trends in Cognitive Sciences* 4 (2000): 197–205; Pylyshyn, "Visual Indexes, Pre-conceptual Objects, and Situated Vision," *Cognition* 50 (2001): 127–158.

34. It is empirically possible for perceptual psychology to have no application. The argument is meant to show that given that perceptual psychology does have application, there are features its explanation must have, by virtue of being perceptual. I take it that the relevant notion perceptual does not simply stipulate these truths. What the argument rules out is singular reference in perception without its being by way of repeatable capacities to discriminate (and represent) aspects of particulars that render them perceivable by the perceptual system. I believe that this is an *apriori* element in perceptual explanation. Excepting occasional conceptual confusion that is inessential to empirically supported findings about perception, empirical explanations in perceptual psychology accord with these points.

35. In principle, I allow for ontological identification of representational perceptual states with sensations or neural states. In principle, I allow for explanatory reduction. Any such reduction would have to explain what representational explanations of perceptual states explain, and thus accord with the second thesis. I am *very sceptical* of such reduction. I do think that perceptual representational states commonly have sensations (conscious or not) as elements, but I take it that *in themselves* sensations do not have any specific representational content.

would produce the same type of sensation or same type of neural disturbance in the brain. Any instance would be equally explanatorily related to an application of a purely schematic perceptual representation (analogous to that, unapplied). To explain what aspects of the particular are purportedly perceived, one must explain the system's response in terms of representational kinds of abilities that are specific to those aspects. Kinds of sensations or neural events *per se* can play a role in explaining nonperceptual, sensory responses to the environment. They obviously play some role even in the full explanation and understanding of perception. But they do not mark acts or abilities that are specific enough to be the perceptual system's ability to refer to (perceive) any definite perceivable entity. The empirical and conceptual bases for the difference between perceptual and mere sensory response is a topic for another occasion.

So to explain a purported (fallibly functioning) *perceptual* context-bound singular representation, perceptual psychology must relate the perceptual response to discriminable aspects of particulars. The perceptual response must be marked as a capacity to represent those aspects. The representational ability is type-identified in terms of aspects of particulars that the system interacts with in successful (veridical) perceptions. Such aspects will help type-identify nonschematic, semantically general, ability general representational kinds that mark conditions of veridicality.

The *second group of remarks on the second thesis* concerns not the general guiding element in perception but the singular element. Representational contents of perceptions and propositional attitudes mark both individuals' perspectival abilities, or exercises of abilities, and the conditions under which the representational function of the perception or attitude is fulfilled. In the exercise of their fallible perceptual abilities, individuals perceive particulars as having general features. The function of perception is to perceive particulars—particular instances of color, shape, or trajectory, particular objects or events—as well as to get their general features right. Fulfillment of this twofold function is veridicality. Perceptions—both the states and the contents that mark them—are veridical when representationally successful. The specification of representational content of a perception should specify a condition that when fulfilled is veridical, not merely veridical-of. It should also specify a content that refers to the particulars perceived. So in specifying the representational content of a perceptual state, one must specify both some element that, when successful, picks out a particular and some element that, when successful, attributes a general feature to that particular.

There are duplicates that would be indiscernible to a perceiver, but where only one of the duplicates is perceived—because it is the one causally responsible for the perceptual state-token, the exercise of the perceptual ability. Or the two might be perceived successfully and

successively, without the perceiver's discriminating the two. Something in the exercise of the perceptual ability must be specific to the particular that is actually perceived. There is no freely repeatable perceptual capacity that distinguishes one particular from another. So a context-bound (ability-particular) element in the exercise of the perceptual ability must be marked in the representational content of the perception. The singular elements are context-bound applications that mark, or are individuated in terms of, actual token exercises of perceptual abilities.

The same train of reasoning applies to the specification of the representational contents of propositional attitudes. Representational contents of attitudes mark both perspectival abilities and exercises of abilities—modes of presentation in thought—and the conditions under which the representational function of the attitude is fulfilled. In the exercise of their mostly fallible propositional attitudes, individuals frequently think thoughts that are about particulars and attribute general features to them. Belief is fundamental to all propositional attitudes, in that all other attitudes are constitutively associated with a capacity for belief. The representational function of belief is to believe truths.

Fulfillment of this function is true belief. Thoughts, both the states and the representational contents that mark them, are true or false. The specification of representational content of a propositional attitude should specify a truth condition—a mostly fallible, partial perspective on a putative subject matter that when fulfilled is true, not merely true-of. Equally, such specification should specify a representational element that indicates the particular that the belief is about. So in specifying a representational content of a propositional attitude that is purportedly about a particular, one must specify both an element that, when successful, picks out a particular and an element that, when successful, attributes a general feature to that particular.³⁶

Parallel considerations to those sketched above about duplicates (or simply considerations about particulars thought about) that cannot be discerned by the individual through freely repeatable abilities demand recognition of singular elements that are context-bound or ability-particular. These elements are applications in thought. Such applications are individuated ultimately in terms of context-bound actual exercises of ability general, syntactically singular, demonstrative and indexical propositional abilities. Specification of such applications is part of the specification of the truth conditions that representational contents of thoughts constitute.

The point that representational contents of propositional attitudes are true or false, not merely true-of or false-of particulars, is further supported by the fact that the representational contents of empirical propositional attitudes are parasitic on the representational contents of perceptions. The

36. For a discussion of representational function, see my "Perceptual Entitlement," section I.

former are individuated partly in terms of the latter. The argument that perception is perspectival and yet picks out particulars seems to me especially evident because of the relative concreteness of our conception of the perspectival and particular-representing nature perception. So the argument regarding perception can be seen as supplementing and to some extent underlying the argument regarding propositional attitudes.

It is hard to overemphasize the significance of these points in evaluating theories of propositional attitudes. Some neo-Russellian accounts, originally developed to account for linguistic meaning, rest satisfied with a propositional "content" that is not fully representational. They incorporate particulars as constituents in the "proposition" that is thought. Such particulars normally do not represent. Physical particulars untouched by minds do not represent. They do not represent themselves or anything else. I find such "propositions" artificial, but otherwise innocuous. But I think that the foregoing considerations show that such approaches cannot provide a full account of perception or thought. A full specification of a perception or a propositional attitude must take it to represent and pick out particulars that it is about. An individual's psychological relation to every element that a psychological state is about is perspectival. The perspective on the particular includes the fallible occurrent *exercise* of an ability general, syntactically singular ability. The exercise is an application. Such exercises themselves are representational acts or events, and are marked by context-bound representations. Thoughts and perceptions are exercises of perspectival abilities that are true (false) or veridical (nonveridical)—not merely true-of or veridical-of particulars. The representational contents of perceptions are veridicality conditions that may or may not be fulfilled by particulars and their properties. The representational contents of thoughts are truth conditions that may or may not be fulfilled by the world.

These reflections on the second thesis supplement the import of the first thesis. Not only is every particular perceived and thought about perspectively. Every particular that is perceived or thought about is represented by elements that are specific to that particular. Such elements need not fully specify or describe the particular. The relevant elements are singular token-individuated representations that mark singular applications of some general representation. The singular application is guided by the general representation. It is specific to the particular not in specifying it, but in being an element in the perceiver or thinker's perspective that is an exercise of a representational ability and that is causally connected in an appropriate way to a particular in the context. Where perceptions and thoughts purport to be about particulars but fail to represent a particular, there is nevertheless a syntactically singular representational element that functions fallibly, and in the instance unsuccessfully, to pick out a particular. This element is of a sort that would have succeeded if its use had been more adept or if conditions had been more favorable.

Let me turn to a *third group of remarks on the second thesis*. Most philosophers who have addressed these topics would accept the second thesis. Most such philosophers impose stronger though in some respects less precise requirements on guidance and attribution than it does. It is common to require that a capacity for perceptual (and perceptually based) singular reference be guided by a large range of conceptual abilities, abilities that are components of propositional attitudes. These have included ordinary sortal concepts; causal concepts; knowledge of one's place in space; veridical, attentive application of most of one's perceptual information in thought; self-conscious application of reasons; and so on. I regard all such requirements as hyper-intellectualized and empirically untenable.

The second thesis does not require that perceptually based representation of particulars be guided by ordinary sortals. In the first place, perceived particulars may include instances of properties and relations, as well as objects and events. In the second, some perceptual reference to physical objects is guided only by rubrics like connected body.³⁷ In fact, the thesis does not require that perception be guided by any *concepts* at all—not sortals, not causal concepts, not concepts of oneself and one's place in space, not quantification or identity. It does not require that most information coming from the referent be veridical. I believe that most of the traditional accounts are mistaken in requiring that perception of physical objects be backed by concepts (components of propositional attitudes) or thought. They conflict with empirical facts about human and much animal vision. They also neglect objectifying elements in perception itself. In being less restrictive, the second thesis may fare better.³⁸

What the thesis requires is modest but substantive. In requiring that perceptually based context-bound singular representation be guided by a nonschematic, general representational ability, the thesis maintains that perceptually based singular representation is grounded in abilities that are independent for their natures of any specific occasions. Perceptually based representation is fundamentally individuated in terms of repeatable kinds. It is embedded in repeatable patterns in the world.

37. For discussion of this generic representational category for physical objects or bodies, see Elizabeth Spelke, "Principles of Object Perception," *Cognitive Science* 14 (1990): 29–56; Fei Xu, "From Lot's Wife to a Pillar of Salt: Evidence That *Physical Object* Is a Sortal Concept," *Mind and Language* 12 (1997): 365–392; Fei Xu and Susan Carey, "Infants' Metaphysics: The Case of Numerical Identity," *Cognitive Psychology* 30 (1996): 111–153; P. W. Juszczyk, S. P. Johnson, E. S. Spelke, and L. J. Kennedy, "Synchronous Change and Perception of Object Unity: Evidence from Adults and Infants," *Cognition* 71 (1999): 257–288; Susan Carey and Fei Xu, "Infants' Knowledge of Objects: Beyond Object Files and Object Tracking," *Cognition* 8 (2001): 179–213.

38. I think that principles governing conditions under which singular application is possible that have been articulated by Quine, Davidson, Strawson, Evans, and others have clear counterexamples. I shall discuss these on other occasions, including a forthcoming book, *Origins of Objectivity*. Cf. my abstract "Perceptual Objectivity," *Kreativität*, XX Deutsche Kongress für Philosophie (September 26–30, 2005), ed. G. Apel (Hamburg: Felix Meiner, 2006), 484.

In requiring that singular perceptual representation be guided by a representational responsiveness to types, the thesis accords with the view that such representation is to be explained—including individuated in anti-individualist fashion—in terms of an ability to discriminate repeatable types (kinds, properties, relations) in the world. Many of these types are not “natural kinds” in the strictest, most universal sense, although perceptual kinds are founded in patterns of the natural world. Some are kinds (colors, for example) that only the representer’s makeup will allow him to be sensitive to. At the level of thought, some types (pianos, for example) have a cultural basis. All have a repeatable character and an independence from particular representational events that makes general explanation of connection-with-the-world possible.

In requiring that singular perceptual representation be guided by semantically general attribution, the thesis attempts to capture the fact that the responsiveness to general types in the world functions primarily to enable representers to represent, and otherwise interact with, *particulars*. The particulars are discriminated only through their repeatable aspects. But the repeatable aspects are of representational significance only inasmuch as they are channels or guides to particulars. This is why at the ground level of perceptually based representation, the guiding representations are *attributive*. They are—when successfully applied—*veridical of* or *true of* particular instances, rather than simply names of types.

I have stated a fairly conservative thesis, and have argued for an even more restricted one. I believe that the second thesis (and the third) can be broadened beyond perceptually based, context-bound singular reference. A broader thesis could include certain sorts of reference to one’s own psychological events and reference to mathematical entities (cf. note 27). I will discuss such cases in section V. But I will not specifically reformulate the second (or third) thesis to deal with them.

The second thesis does not require that the guiding representation be *veridical of* the referent (if any) of the singular representation. The thesis sets a condition on perception and all perceptually based thought that purports (functions fallibly) to refer to particulars. I turn now to necessary conditions on *successful*, perceptually based, context-bound singular representation.

III

All reference in perception or thought must occur within a topological perceptual structure, or a logical-inferential propositional structure, that type-identifies a network of representational abilities. These structures are fundamental to psychological explanation as well as to a reasonable epistemology.

The second thesis claimed that an aspect of such structures is a constitutive connection between perceptually based singular reference and attribution. The third thesis presupposes the second, but it makes a claim not about psychological structure and psychological ability but about conditions for successful reference. Intuitively, the key claim is that, in each case, successful perceptually based singular reference can occur only if some guiding attributive is *veridical of* or accurately characterizes the referent.

More specifically, the third thesis is:

Laying aside applications of egocentric indexes, if an autonomously used, perceptually based, context-bound singular representation is to have a referent in perception, perceptual memory, perceptually guided actional state, perceptually grounded intermodal state, or propositional thought, the singular representation must be guided by some empirically committal, nonschematic, attributive, semantically general, ability general representation that is in fact veridical of the referent.

The thesis maintains a minimum role for *veridical general representation* in determining successful reference by perceptually based singular representation.

Recall that a *perceptually based* representation is one that is or occurs in a perception, or that occurs in perceptual memory that purports to represent a previously perceived particular, or that occurs in a perceptually guided actional state, or intermodal non-propositional perceptually grounded state, or that is a thought or thought component that functions to represent a particular through perceptual resources.

To be *empirically committal* is for a representation to depend for its content and the warrant for its application on perception. So is self-identical, is the result of adding 7 to 5, and is the result of adding 7 red squares and 5 red squares are not empirically committal. The thesis holds that for perceptually based, context-bound, singular reference to be successful, the singular element must be guided by an empirically committal general representation that is *veridical of* the referent.

Work by Kripke and Donnellan showed that uses of context-dependent singular terms can succeed in having a referent even if salient descriptions associated with them fail to be true of the referent. Both authors cited cases that are easily extrapolated to show that no explicitly applied linguistic predicate must be true of the referent.³⁹ One can refer to the man in the corner who pretends to drink what is in fact a martini even if one thinks of him as the woman along the wall drinking a soda. One can

39. Saul Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1972); Keith Donnellan, "Reference and Definite Descriptions," *Philosophical Review* 75 (1966): 281–304; Donnellan, "Proper Names and Identifying Descriptions," *Synthese* 21 (1970): 335–358. Regarding extrapolation, I have in mind Donnellan's case involving the child in his 1970 paper and Kripke's Jonah case.

refer to Socrates even though one calls him "Hebrides" and is mistaken about what he is known for.

A similar phenomenon occurs in perception, independently of language. One can see a white sphere behind one—where one's sight is guided by a prism one is unaware of—even if one perceives it as a brown animal in front of one.

Such cases show that reference can succeed even though the salient general representations that accompany a context-bound singular application are not veridical of the referent. I take for granted that for context-bound singular representations, no set of general representations in the repertoire of the individual need uniquely fix the referent of a context-bound singular representation by being true of or veridical of that referent.⁴⁰ In many cases, no general representations in the individual's repertoire *can* uniquely fix the referent by being veridical of it. This background point applies both to singular reference in thought and to singular perceptual representation.

Given their focus on language, Kripke and Donnellan did not address what I regard as a deeper, representationally prior issue. They did not inquire into the background of mental representation that is inevitably present in such cases. They did not ask whether the cognitive subject must presuppose or apply some empirically committal, general representation in perception or empirical thought that is veridical of the referent, if a context-bound singular representation is to succeed in referring.⁴¹ Can singular representation succeed, even though *no* guiding general representation applies veridically to the referent? The third thesis maintains a negative answer. Empirical *de re* reference requires empirical attribution that is veridical of the referent.

Arguably, *through interlocution* an individual can refer to an object, even though literally no general representation that the individual is disposed to apply is true of it. Two adults may call a warp in space-time "Sam" and tell a child some falsehoods about Sam. The child accepts the stories and builds a fantasy life about Sam. The child has no meta-concepts. The child cannot think of Sam as the object the adults were talking about, or as the referent of the name. The child might garble the name and think of the object as Slam. The child might lack the super-

40. Cf. Donnellan, "Reference and Definite Descriptions"; Donnellan, "Proper Names and Identifying Descriptions"; Kripke, *Naming and Necessity*; David Kaplan, "Demonstratives," in *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein (Oxford: Oxford University Press, 1989); and my "Belief *De Re*." I think that this is a necessary truth about empirical thought. Cf. "Belief *De Re*." Certain types of singular reference ineliminably depend for being successful on causal relations external to the thinking individual.

41. I am interested in reference in perception or thought. I think a counterpart thesis for linguistic singular reference may be true. The rough counterpart for linguistic reference would be: the individual must apply or presuppose in perception or thought some nontrivial semantically general representation that is true of the referent of an autonomously used, singular, context-bound *linguistic* representation, if that singular representation is to have a referent.

ordinate concept of a spatial object (and the adults could have named a number). No guiding general representation true of the referent is available to the child. Some will deny that the child thinks about anything. I see no sound theoretical basis for such denial.⁴²

Such cases are very special. They depend on special features of interlocution and on the linguistic institution of proper names. The child's use of the name in thought picks up a reference grounded in others' uses. I believe that an individual could not think about an object *autonomously* if he or she had no more kind-typing capacity with respect to the relevant object than the child has. At any rate, that is an element in the view that I want to explore. Since the child's ability to refer with the name in the absence of successful categorization seems to derive from the anaphoric or parasitic character of the child's thought, I lay interlocution aside.

I shall discuss the third thesis by centering on singular reference in perception.

It is clear that an individual and his perceptual system can mistake nearly all features of an entity, yet still succeed in perceiving (and thinking about) the entity. Color, texture, shape, surface properties, spatial location, size, motion, and sortal type can be misperceived—all in a given instance. Yet the individual, animal or person, can still see an object that is appropriately causally related to the perceptual representation. For example, one can see an entity through an unknown prismatic distortion and have so limited a view that one gets its shape, color, location, and sortal type wrong. One can see something as an object with a definite surface, whereas it is in fact a coherently formed wisp of fog or a strikingly salient beam of light, perhaps a hologram. One can form mistaken beliefs about *that* entity.

There are, however, limits on how mistaken a perception can be while still having a perceptual referent. I conjecture, for example, that to be visually perceived, an object must produce a representation that discriminates it from the surround and represents it correctly as having an approximately bounded shape that is in fact trackable in generic form. The specific shape might be misperceived. But if the object is seen, its

42. For such a denial, see Gareth Evans, *The Varieties of Reference* (Oxford: Clarendon Press, 1982), 105–120. Evans requires that to refer to an object, one must know which object it is by discriminating it from other objects by perception, description, or recognition. More generally, he requires that one know what sort of thing would make one's thought true. In working out what this requirement means, he places restrictions on reference through ordinary cognitive capacities that I regard as poorly motivated and quite unacceptable. I find his arguments for this view as applied to memory and interlocution (127ff.) unpersuasive, his account of thoughts about natural kinds (e.g., 117) mistaken, and his strictures on perceptual belief itself excessive (151–170). For criticism of Evans that I largely agree with, see Marleen Rozemond, "Evans on *De Re* Thought," *Philosophia* 22 (1994): 275–298. I think that his view is a rearguard defense of the old over-reliance on individual knowledge and control in determining a referent, an over-reliance driven by the traditional philosophical hyper-intellectualization of accounts of thought.

shape must bear some systematic relation to the specific presented shape (such as being a bounded deformation of it).

Suppose that light comes from an odd angle. Its reflection off particles in the air causes a representation as of an object as straight ahead. Suppose that neither the light nor the particles form any coherent, trackable shape analogous to the apparently trackable shape of the apparent object. The light is not a flash with a shape that is a deformation of the shape represented, nor is it a hologram. Then I think neither the light nor the dispersed particles are perceptual referents, with misperceived features. They are not seen at all. There is only perceptual referential illusion.

Perception here does not fail merely because the light is not where the perceptual system represents an object as being. We can perceive things while mislocating them. Perception does not fail because we cannot see light or dust and mistake them for more mundane objects. We can.⁴³ The problem is that the light and particles lack any spatial coherence that is like the bodies and surfaces normally tracked by the individual or his visual system. The visual system's binding various representations together into a representation of a single entity does not correspond to any such system of properties in the environmental causes of the complex representation. The seen entity, if it is seen as a body, must have something like the boundedness of a trackable object. Perceptual systems have abilities to group and track such generic shapes as *bounded*. They therefore have such generic representations.

The requirement is certainly not that perceived entities must be internally spatially connected. We see constellations, flocks of birds, and so on. In some of these cases, we also see several bounded objects at once, and successful perceptual reference to the group depends on seeing a sufficient number of the component objects as bounded. In other cases, the gestalt of (approximate) boundedness of the whole is all that matters. I will not try to specify here the exact identity of representations that

43. Thus it is important to see here that the veridical guiding perceptual representation need not be the ecologically or practically most fundamental guiding representation, the basic categorizing attributive—in cases where one perceives an object. One can perceptually represent a particular as a body with an integral connected shape, and be mistaken about the particular's being a body. What makes perceptual reference possible—what makes it possible that one sees the particular flash of light or hologram—is the guiding, generic shape representation, not the guiding body representation.

It may be true that to perceive physical bodies, one must have and sometimes attribute the guiding perceptual representation body. It does not follow that when it is attributed, this representation must be veridical of a particular if the particular is to be seen. It does not even follow that one cannot, on occasion, perceive a particular physical body in cases where one does not attribute body to the particular. One might perceive something as a flash of light with a certain connected form, and what one actually perceives is a body with that form. I believe that in the case of *successful perceptual reference* in vision, shape attributives may be more fundamental than the attributive body, even though the latter has a more fundamental role in the explanation of much action.

must be veridical if various types of perceptual references are to succeed in being veridical. But it seems to me that these examples are suggestive.

Argument for a Restricted Version of the Third Thesis

I would like to give an argument that explains the truth of the third thesis. I shall again concentrate on specifically *perceptual* representation.

Suppose that a perceptual, context-bound, singular representation *S* picks out some particular entity *E*. *E* must be of some kind and must have properties. For *E* to be perceived, some aspects or properties of *E* must play a role in the causation of *S*. These aspects or properties must be one of a group of causal factors that are instances of a pattern.⁴⁴ If stimulation of the sensory organs were to occur within the parameters of the pattern again, another singular representation would be triggered. If the distal stimulation were of the same type, and it produced relevantly similar proximal stimulation, the new singular representation would also be successful. There is room, of course, for various types of stimulation to trigger any given singular representation. But this variety must fall within some general repeatable pattern if the success of any given occurrence of singular representation is to constitute exercise of an ability—and if success is to be explainable.

E's causal contribution to the general pattern must be through some of its properties or aspects. Some of *E*'s properties or aspects will be irrelevant to producing the pattern. If *E* is a body, properties of the sides of *E* that do not reflect light to a visual system will make no contribution. Or if *E* is an instance of shape property, certain aspects of the instance (its angles' adding to 180 degrees, or its touching a circular entity) may make no contribution. Call the set of relevant properties "*P*."⁴⁵

Some proper subset of *P* is of representational significance. For example, indiscernible components of *E* may be causally relevant but representationally irrelevant. The aspects that are representationally relevant are those that the perceptual system can discriminate and represent. A condition on perceiving a particular—on applying a perceptual context-bound singular representation to it—is that the individual or his perceptual system perceptually discriminate the particular by percep-

44. I am not suggesting that it is metaphysically necessary that any causation be part of a pattern. Rather, given that we are dealing with perception (an empirically determinable matter), it is necessary that perception be an ability that is responsive to certain patterns of stimulation. Such stimulation will produce perceptual representations given any stimulation within the pattern, if the antecedent psychological states of the perceiver are fixed. This macro (approximate) determinism is well entrenched in any empirically reasonable theory of perception.

45. If *E* is an instance of a relation, matters may be more complex. There are issues about perception of relations that I have not sufficiently understood, much less explained.

tually discriminating some of its properties or aspects.⁴⁶ The point grounds psychological explanation of perception. For it to be psychologically explainable how the individual or perceptual system discriminates E, some of E's properties that help cause S, some subset of P, must ground an explanation of how the perceptual system perceptually discriminates the particular E. There must be something about E that enables the individual and perceptual system to discriminate it perceptually from other particulars, possibly feature-instances, that causally affect the perceptual system, but that are representationally irrelevant.

By the argument for the second thesis, perceptual responses by the perceptual system must be type-identified in terms of a nonschematic, attributive, semantically general, ability general representation that guides S. Perceptual discrimination of E, involving successful singular representation of E, is a perceptual response. So perceptual discrimination of E via a subset of P (a subset of E's properties) must be explained in terms of a perceptual response type-identified by such general representations. For an individual or perceptual system to perceptually discriminate E via some subset of E's properties in a way that is systematically and correctly explainable in terms of a representational response by the perceptual system to the causal effect of those properties—a response type-identified by a nonschematic, semantically general, ability general attributive—is for the individual or perceptual system to apply such representation veridically to E.

If the individual or perceptual system does not perceptually discriminate the particular (even partly) in terms of aspects of the particular, there could be no explanation of perceptual discrimination of that particular rather than any number of other particulars that figure in causing the same effects on the perceptual system. More fundamentally, there could be no way that the perceptual system representationally discriminated the particular. The problem of accounting for representational specificity again lies close to the heart of the argument.

The argument applies no matter what E is. As I have emphasized, many particulars enter into causation of a successful singular representation in perception. Particulars in the sensory pathways or in the causal chain that links perceptual objects and proximal stimulation can be excluded on the ground that they are not at the right ecological "level"

46. The distinction between individual and perceptual system is not, I believe, significant for this argument. Many transformations on representations carried out in the perceptual system are not consciously available to, and are not acts by, the individual. Similarly, the fundamental principles governing the formation and transformation of perceptual representations are modular and not available to the individual. On the other hand, most of the perceptual representations in the perceptual system—certainly the conscious ones—are also perceptions by the individual. The point of explanation in perceptual psychology is to explain the perceptions that individuals have, and the representational successes and failures of these perceptions. The explanation locates those perceptions (perceptual representations) in a perceptual system governed by general principles.

to enter into explanations that bear on perception and action on the environment. Individuation of perceptual representation is partly governed by such explanations. There will still be various particulars that figure in causing any given occurrence of a context-bound singular perceptual representation *S*, but that do not figure in the psychologically relevant explanation of the *perceptual* discrimination of *E*, the referent of *S*. They are not particulars that the perceptual system responds to in discriminating *E*. To successfully perceive *E*, the individual and system must get something about *E* right. They must exercise a perceptual ability to discriminate some aspect of *E* that distinguishes *E* from other particulars (a) that figure in explanations of the individual's use of perception in satisfying needs; and (b) that play a causal role in the production of *S*.

The argument implies that even if salient perceptual representations fail to be veridical of a perceived entity, some guiding representation must be veridical of it if there is to be a psychological explanation of how the perceiver or his system discriminates that entity in the context. If place, kind, color, and particular shape are misperceived, there must remain something about the entity that the perceiver and system get right. Getting something right discriminates the entity by guiding the singular representation through a feature of a particular that distinguishes the particular from particulars of other kinds in the environment that figure in the causation of the particular perception.

This point seems to me just as applicable to perception of features as to perception of individual objects or events. One might see some green but see it as yellow; one might get its position wrong; one might get wrong what object it is a color of. But one must discriminate it in some way. Perhaps one must get right that it is a more or less connected expanse of a more or less uniform color. If even such generic representations as this one go wrong, there is no seeing a color instance. There is no explanation of wherein the perception is of the color rather than of an associated shape, or some part of the color expanse, or the surface.

I want to make a more abstract point about what must be "gotten right" if an entity in the environment is to be perceived. The visual system represents entities as located outside itself. Doing so is central to the role of perception in generating motor activity geared to coping with perceived entities. We can perceive entities even though we are mistaken about where they are. But if the visual system represents an entity as being in a location and the representation is caused merely by some event in the optic nerve, nothing is perceived. Perception fails because the causes of the representation are not located in the environment. The system's commitment to the entity's being located in the environment must be veridical if normal visual reference is to succeed.

Not understanding this point lies at the root of the hoary mistake that we see sense data or that we see hallucinations. One can understand the phenomenal basis for such usage. There is a sense in which it is harmless

as (I think stretched) ordinary language. But a philosophy that does not distinguish a notion of perception and sight that firmly excludes such cases as instances of successful perception misses out on fundamental functional and individuating distinctions that lie at the basis of understanding the nature of perception and perceptual representational kinds.

It does not follow from *this* point that there must be a veridical perceptual representation that accompanies and guides successful perceptual reference. In particular, it seems implausible to presume that the perceptual system has an abstract representation like spatially located in the external environment. Although such general *conceptual* representations have contrast value for a mature thinker, they do not provide any usable distinction for a perceptual system. So located in the external environment is not a perceptual attributive, a representation available to the perceptual system itself. Subhuman primates and young children probably lack any counterpart *concept*, even though they incorporate their perceptual representations into a belief system.

So the abstract point is a weaker, less committal point than the third thesis, or the argument for it. I think it worth articulating because it is relevant to the epistemology of informed reflection.

Determining what representations must be veridical if various particulars are to be seen is a complex empirical matter. There is information even in nonveridical kind-attributions from which one can, on mature reflection, abstract general *concepts*, like located in the external environment, that must be *presupposed* to be veridical of perceived particulars. One can do this in the absence of knowledge of what general representations *used by the system* must be veridical of perceived particulars. Such concepts are presupposed by perceptual systems.

Let me shift gears. We can imagine a sophisticated adult in disorienting circumstances thinking: "I do not care whether *that* is spatially located in the environment in the usual way. It may be a reflection on the retina. Or it may be an internal image. I want to know what *that* is."

It does seem possible to refer in this way.⁴⁷ Is the reference in thought unaccompanied by any general representation that is true of the referent? I believe that carrying out such a reference involves canceling the normal commitment to an environmental location for the object of perceptual reference. To do this, the thinker must *have* concepts of appearance, perceptual representation, retinal image, and so on—in addition to physical object concepts. An individual that knew nothing of reflections on retinas or perceptual representational images could not make reference to them. Thus the thinker *has and applies* a disjunctive concept that guides disjunctively and that is true of the referent: either physical object or retinal image or perceptual representation.

47. The sophisticated adult understands the demonstrative that to apply, in a default manner, to a physical object if the perceptual system is successful in perceiving such an object, and to a retinal or (presumably as third backup choice) internal image if it is not successful.

Children and apes probably cannot think of their perceptions as being caused by objects. Neither they nor their perceptual systems represent causal relations between perceived objects and perceptual representations.⁴⁸ But an individual that thinks the thought just discussed probably must do so. Taking retinal reflections and internal images as possible referents of a perceptually guided demonstrative requires having some concept of possible causes or explanations for perceptual representations other than the normal ones. I think that such an individual presumes a meta-view of the referent as a cause of the perceptual representation.

I think that appeals to such meta-representations in accounts of reference are usually cheap (explanatorily unilluminating) and mistaken. In this case, meta-representation is forced by the individual's sophisticated move of canceling the standard presumption of spatial location generated by the function of the perceptual system. Even here, the meta-representation is not sufficient to fully specify the referent. The individual is usually incapable of specifying which relevant cause or explanatory factor is the referent. The general meta-representation provides only a loose restriction on the reference, at most a necessary condition, not a sufficient one. The demonstrative applicational element in the thought remains irreducible.

The limits of perceptual error form a rich topic. The third thesis invites more exploration.

IV

The first two theses concern conditions on the role of psychological abilities in determining perceptually based representation. The third thesis is about conditions on successful, context-bound, singular, perceptually based reference. The fourth thesis is about opportunities for knowledge implicit in the psychological and referential requirements so far discussed.

The anti-individualist role of contextual and causal relations in fixing *de re* reference, and in individuating the representational content of psychological states, has seemed to some philosophers to displace the definitional and inferential relations among concepts that were traditionally relied upon as bases for claims of apriori knowability. It is true that

48. The view that causation is a perceptual category that applies in perception to the relation between perceptual objects and perceptions themselves was defended as a thesis in psychology some years ago. Cf. A. Michotte, *The Perception of Causality*, trans. T. R. Miles and E. Miles (London: Methuen, 1963; translation of 1946 French edition). There are numerous empirical objections to the theory, and it is no longer taken seriously in psychology. (Some psychologists take seriously Michotte's less ambitious but still controversial view that *perception* attributes causation between perceived physical events.) John Searle argues for the stronger Michotte view in *Intentionality* (Cambridge: Cambridge University Press, 1983). I criticize his argument in "Vision and Intentional Content."

not only context-bound singular reference but also semantical applicability of empirical, semantically general, ability general concepts and percepts are almost never fixed by definitional or inferential relations among concepts. Such reference and semantical applicability are partly dependent on nonrepresentational (for example, causal) relations to the environment. In view of this fact, one might conjecture that nothing can be known apriori from singular *de re* perceptually based reference, and that nothing can be known apriori in using empirical concepts—beyond logical and mathematical truths involving such reference or such concepts inessentially. One might deny that one can know apriori such conditionals as if that [perceptually presented as a body] is anything, it is a dog, or if something is a cat then it is an animal. One could generalize to a view according to which empirical concepts are introduced to apply to *whatever best empirically explains their introduction*, where it is a very open empirical question how to explain any given introduction.⁴⁹

I will not try to adjudicate the examples just mentioned. I think, however, that the general view is mistaken. The fourth thesis is:

Some of our perceptually based de re states and attitudes, involving context-bound singular representations, can yield apriori warranted beliefs that are not parasitic on purely logical or mathematical truths.

The apriori warrants depend on the *nature* of perceptual representation.

I will not try to specify exactly what beliefs count as parasitic on logical and mathematical truths. An example is: if that entity exists, it is self-identical.⁵⁰

49. This is a form of indexicalism. For discussion of other forms of indexicalism, see my "Phenomenality and Reference: Reply to Loar" and "The Indexical Strategy: Reply to Owens," in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, ed. Martin Hahn and Bjorn Ramberg (Cambridge, Mass.: MIT Press, 2003). Holding that all our empirical concepts are indexical misses the evident specificity of our conceptual and perceptual representations.

I think it beyond serious doubt that our ordinary empirical concepts are not introduced in the way indicated in the text. But would concepts so introduced fail to support the fourth thesis? The apriori connection between the concept and the concept of an explanation would remain. The introduction, associated as it is with perception, must presume that what the concept applies to has causal properties. If the concept is allowed to depend on the perceptual representations of the introduction, there is the further connection to the concept of spatial location, and any other concepts that conceptualize limitations on the relevant perceptual reference. Thorough discussion would require investigating conditions on what counts as an introduction of a concept. These are the perceptual and presupposed unifying conditions that enable an explanation to get started.

50. Some believe that use of the demonstrative either guarantees a referent or guarantees commitment to there being a referent. I think it clear that one can use a demonstrative while being aware of the possibility that one could fail to refer. Suppose that one is in a psychological experiment where one has already learned that some fraction of

I think that there are apriori knowable truths derived from basic limitative principles of the sorts indicated by the third thesis. These principles govern reference of perceptual representations in perceptual systems. For example, some truth like if that object [visually presented as a body] is any object at all, it has a trackable, integral three-dimensional form is apriori knowable. Further, if that object [visually presented as a body] exists, then it is spatially located in the environment is apriori knowable. Suppose that the individual does *not* cancel the default presumptions of perceptual judgments. There is no special allowance for reference to retinal images or internal perceptual representations. In such cases, no sense experience need figure in a warrant for the relevant beliefs. Belief in such truths can be warranted by reflection on limitative principles governing perceptual reference. Thus apriori knowledge and apriori epistemic warrant can exploit limitative principles deriving from the third thesis.

These truths are nonlogical, nonmathematical apriori truths (apriori knowable truths) about perceived objects. I shall first discuss the sense in which they are nonlogical or synthetic. This sense will suggest why they are nonmathematical as well. Then I will develop the sense in which they are apriori.

It may appear that the conditional truths associated with the limitative principles are tantamount to logical truths of the form if something is a brown oblong object, it is brown and oblong or if something is a body with an integral three-dimensional form, it has an integral three-dimensional form. The appearance is deceptive.

Perceptually based thoughts are rather special. When perceptually guided in thought, the demonstrative *that* can succeed in referring even if some of its accompanying general representations are not veridical of its referent.⁵¹ Thus in thinking the perceptually based thought that brown oblong body is brown and oblong, one can perceive and think about a body referred to by the token-application of *that* even though the body is not brown or oblong. In these cases, one cannot infer from the success of one's perception of an object and the nature of the perceptual attribution (or perceptually based attribution in thought) *alone* that the perceived object has the perceptually attributed properties.

In the third thesis, I held that perceptually based referential success is *not* possible if *certain* general representations are not veridical of the referent of the context-bound, singular element. I held that is spatially located must be true of objects of visual perception, in the absence of an

one's attempted perceptual or other demonstrative applications have failed. The example discussed near the end of section III points in this direction. I believe that perceptual demonstratives are not infallibly successful in having a referent, because perception is not infallible. I believe that in given cases one is quite able to realize this and to suspend belief about whether a given reference has succeeded. The examples that follow assume these points.

51. Cf. Donnellan, "Reference and Definite Descriptions."

explicit cancellation of a default presumption that underlies visual perception. I conjectured that some general representation of trackable, more or less connected, or integral, spatial form must be veridical of the referent of context-bound singular perceptual representations that pick out physical bodies. I conjectured that some color attributive must be true of any color instance that is perceived. And so on.

The points made in the preceding two paragraphs indicate that there are important differences among perceptually based attributives in their roles in purporting to restrict perception of particulars. Some perceptual attributives that guide singular elements in perception, or that are presupposed by perception, must be veridical of the perceived particular if perceptual reference is to succeed. Other perceptual attributives accompanying the singular elements need not be veridical of the referent.

For example, with respect to a purported perceptually based reference to a body, (something like) integral three-dimensional form must be veridically attributed in perception to the perceived body if perceptually based reference to a body is to succeed. Similarly, for conceptualizations of this perceptual attributive. The conceptual attributive spatially located presupposed in visual perception must also be true of a visually perceived referent if the referent is to be perceived (assuming no cancellation of the default position). By contrast, brown and oblong need not be veridical when they are attributed and the body is perceived. Similarly, perhaps more or less uniformly colored must be veridical of a perceived color instance, if a perceptually based reference to a color instance (as of orange, for example) is to succeed. By contrast, orange need not be veridical of the perceived color instance.

How do these points bear on knowledge?

There is, I think, a certain natural construal of

(LT) If that brown oblong object exists, it is brown and oblong

(LT') If that object is a trackable, integral three-dimensional body, then it is trackable, integral, three-dimensional

on which the relevantly expressed thoughts are logical truths. They can be known through understanding their forms. On this construal, the antecedents cannot be true unless there exists a demonstrated object *and* it has the properties attributed to it (e.g., brown and oblong). Then, of course, if the antecedent is true, the consequent must be true. The truth hinges on logical form.

There are closely related thoughts, however, that are not logical truths. These thoughts can perhaps be elicited by this formulation:

(EC) If that object [perceptually presented as a brown, oblong body] exists, then it is brown and oblong.

I take it that the thought relevantly expressed by this formulation can be known to be true only if either the perceived object is *empirically* known to be brown and oblong, or there is some other empirical knowledge that

any object perceived in the context will be brown and oblong. A connection between the perceived entity's existing and its being brown and oblong can be known only empirically.

By contrast, consider

(AC) If that object [perceptually presented as a trackable, integral, three-dimensional body] exists, it is trackable, integral, three-dimensional.

(AC') If that particular [perceptually presented as an instance of orange] exists, then it is a more or less uniform expanse of some color.

(AC'') If that particular [presupposed in visual perception to be spatially located, where the default presumptions are not canceled] exists, then it is spatially located.⁵²

Here the connection between successful perception and the guiding attributive (in the first two examples) or the presupposed attributive (in the third example) can be known apriori.⁵³ Successful perception itself is dependent on the attributive's being veridical of the perceived object. Perception must figure in the warrant for a belief that the perceived entity exists. It must contribute to the warrant for believing that there is an entity with the attributed feature. But epistemically, the existence of the perceptually referred-to entity and the veridicality of the basic attributive as applied to the entity go together. Warrant for belief in the conditional comes as a nonempirical package.

The *connections* between the entities' being perceived to exist and their having the relevant attributed features can be known nonempirically in (AC) through (AC''). This is because the connections are epistemically relevant conditions on successful perception. Being warranted in relying

52. The resemblance between the notation in these formulas and David Kaplan's "dthat" notation is intentional. Cf. his "Demonstratives," in *Themes from Kaplan*, ed. J. Almog, J. Perry, and H. Wettstein (New York: Oxford University Press, 1989), section XII. The main differences are (a) that here we are discussing mental representation, not language, (b) that here what goes in the brackets is an attributive, not a singular term; (c) that here what goes in the brackets expresses attributives that are attributed or presupposed in thought by the individual thinker. I take it that the demonstrative can succeed in referring (in thought) to a particular even though some of the attributives applied in thought and expressed within brackets fail to be veridical of the particular. The point of the present discussion is that in such cases, there must be some attributive associated with the application of the demonstrative that is veridical of any particular that is referred to. I believe that certain guiding attributives have this status for given types of *de re* reference.

53. There are interesting issues in distinguishing the role and epistemic status of different types of restricters that can be known apriori to be necessarily connected to successful perceptual reference. For example, I think it important to distinguish restricter attributives like integral three-dimensional form from those like trackable. Unlike attributives that apply to generic spatial forms, trackable is a representation that is not likely to be attributed to perceived objects in perception itself. A body must be trackable if it is to be perceived. I think that we can know this apriori through reflection. But attributives that apply to the relation between perceived entities and the capabilities of the perceptual system are not employed by the perceptual system. Thus trackable is like spatially located in being pre-

on the connection between the perceptually based attributives and the application of the perceptually based demonstratives is a condition on being warranted in applying the perceptually based demonstratives. So warrant for perceptual applications—for perceptually warranted perceptually based thought—presupposes warranted connection between those applications and the relevant guiding (or presupposed) attributions. The connections derive from conditions on veridical perception itself.⁵⁴ So the empirical warrant involved in perceptually based thought—warrant that derives from perception—presupposes a warrant for accepting the relevant connections. So the warrant for accepting the relevant connections is not empirical. Since thoughts of the sort expressed in (AC) through (AC'') make commitments to the relevant connection, belief in them is warranted nonempirically.

Thus the basic limitative principles on perceptual reference yield apriori knowledge. The thoughts that connect the referents of perceptually based demonstratives with the properties attributed by the basic attributive restricters on singular perceptual reference can be known apriori. In these cases, one *can* infer apriori from any success of one's perception of an object and the nature of the perceptual attribution (or perceptually based attribution in thought) that the perceived object has the perceptually attributed properties.

It should be carefully noted that the apriori warrants for instances of fundamental limitative principles governing perception do not derive purely from understanding the logical form of the relevant thoughts. Thoughts expressed by (AC) through (AC'') are knowable apriori, but the thought expressed by (EC) is knowable only empirically. What lies at the root of the warrants for the thoughts expressed by (AC) through (AC'') is reflection on the nature of perceptual reference, not on general logical principles. In this sense, truths like those expressed by (AC) through (AC''), and the limitative principles underlying them, are not logical truths. Knowledge of them does not depend purely on understanding the form of the thought. It depends also on reflection on the particular perceptually based predicative concepts involved, and on the limits on perceptual reference. The most general principles limiting perceptual reference are associated with nonlogical, nonmathematical truths that we can know with apriori warrant.

The guiding representations underlying successful reference seem to be much more generic, at least in perception but I think also in thought, than the tradition commonly supposed. Perception of individuals and empirical thought about them can survive much more fundamental

supposed by the perceptual system, although I think that the ground of the presupposition is different in the two cases.

54. I think that belief in connections of this sort is warranted because reflection shows that each such connection is a general, representational, necessary condition on the success of the relevant type of perceptual representation. Warrant presupposes reliable success in

error than many theories allow. Still, the limitations on successful empirical reference are not warranted through perception. They are conditions on the possibility of perceptual experience and empirical thought about the world.

The epistemology of restrictions on perceptual reference is a rich topic. I want to comment briefly on one complex issue here: It is important to distinguish apriority from availability to armchair reflection.

I shall discuss cases in which some apriori knowledge is not armchair knowledge but depends for its *availability* (as distinguished from its warrant) on specialized empirical investigation. I take it as obvious that some armchair knowledge is not apriori knowledge. By epistemically relying on widely available empirical truths, one can reason to conclusions without engaging in empirical investigation.

I believe that *some* of the apriori warranted knowledge that resides in the limitative principles is available to such armchair reflection. The general arguments for the existence of attributional restrictions, which I gave in sections II and III, are both apriori and available to armchair reflection. Moreover, some of the specific knowledge that instantiates those principles is both armchair and apriori. For example, the requirement that perceived bodies be trackable over at least short periods of time (even if they are not in motion) seems to me to derive from reflection on the need to distinguish body representations both from event representations and from the key shape representations that guide (visual) body representations.

On the other hand, some apriori knowledge that resides in the limitative principles can be elicited only through empirical investigation. This is not to say, of course, that the knowledge is *warranted* through empirical investigation. It is to say that it is *available* only through empirical investigation. Even armchair apriori knowledge is probably available only through widely shared empirical experience, which gives one the conceptual wherewithal to recognize apriori truths. Some apriori knowledge can be elicited only through empirical experience that is more specialized than experience that is easily available to everyone. The empirical investigation helps determine what representation plays the role of making successful singular perceptual reference possible. It justifies a judgment about the identity of a condition-setting representation. It does not justify the role of the representation in setting the conditions on successful empirical perception. That function is not, and as far as I can see cannot be, warranted empirically.

This situation is implicit in the difficulty we have in specifying *exactly* what generic spatial form is used by human visual systems as a minimum guide in seeing physical bodies. What, more specifically, is the topology gestured at by the terms that we have been using—"integral" or "more or

appropriate conditions. So certain very general conditions on success are also conditions on warrant. These issues need, of course, further exploration.

less connected," or the restriction on color reference that I gestured at in the phrase "more or less uniform"? Indeed, recognition that body is a generic perceptual category, in some ways more basic perceptually than categories for characteristic shapes of specific natural kinds (the shape of a duck or a tree), was the result of empirical investigation (cf. note 37). So empirical investigation is important in identifying the relevant representations used by perceptual systems, and in recognizing their centrality. Such investigation is often necessary to the formulation of specific limitative principles that can be known apriori. But the warrant for the connection between the basic attributive restricters and successful singular reference in perception (hence in some perceptually based thought) is ultimately apriori.

Are there further apriori truths of categorization, beyond those associated with principles governing perceptual reference? Are there apriori truths about the range of application of an empirical concept that are not mere instances of mathematical or logical principles and that do not derive from principles governing the referential limits of a perceptual system? Putnam's claim that cats are, if anything, animals is not apriori points in the direction of a negative answer.⁵⁵

I conjecture a positive answer. I think that most empirical concepts are associated with superordinate concepts that provide necessary conditions for their application. I think that one can be defeasibly apriori warranted in believing general limitative principles governing the range of application of concepts. For example, we can know apriori that water is, if anything, physical and occupies space; that if something is yellow, it is colored; and that cats are if anything entities with physical properties.⁵⁶

Despite my advocacy of the fourth thesis, I believe that the kinds of connections that are apriori are, for the most part, very generic. The point that we know only empirically that a kind like gold or water is a natural kind and has a unifying empirical principle was already made by Kant. The taxonomic arrangement of genus and species is vastly more fluid and empirically sensitive than it seemed to be two centuries ago. The apriori connections of the classificational sort that I have discussed here are mostly between relevant concepts and superordinate concepts for *very* generic features or relations. Exploring apriori limits on empirical representation promises to be both difficult and rewarding.

55. Hilary Putnam, "It Ain't Necessarily So," *Journal of Philosophy* 59 (1962): 658–671; reprinted in Putnam's *Philosophical Papers*, vol. 1 (Cambridge: Cambridge University Press, 1975). See also Putnam's "The Meaning of 'Meaning,'" *Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press, 1975).

56. It is important to remember here that being apriori warranted is not equivalent to being invulnerable to empirical counter-considerations. Apriori warrant concerns the source of positive support, not sources of possible overthrow. A belief can be apriori warranted even though it is vulnerable to possible empirical overthrow. The mere fact that it is epistemically possible that it turn out that there is no space or time does not show that our

V

The *de re-de dicto* distinction reaches far back into philosophical tradition. Much of this tradition concerns modality. Some of it concerns representational states. Modern discussion of *de re* states and attitudes stems from reconsidering Russell's notion of acquaintance.⁵⁷ In "Quantifiers and Propositional Attitudes" (1955), Quine made a show of reviving the distinction in his analysis of belief sentences. Although he was sensitive to the intuitive epistemic distinction, his philosophical concerns were almost entirely logical and linguistic.⁵⁸

In "Quantifying In" (1969), Kaplan turned discussion back in what I think to be the right direction—toward understanding the epistemology of attitudes, not the linguistic form of attitude *attribution*. He explored an "*en rapport*" representational relation between the cognitive subject's beliefs and some *re*.⁵⁹ Kaplan avoided Russell's untenable epistemology and philosophy of mind. He sought a relation underlying Russell's intuitions but grounded in everyday considerations. Kaplan was guided both by *linguistic phenomena* associated with quantification into contexts of belief attribution and by *cognitive paradigms* of perception and perceptual memory. I think that he leaned too much on the linguistic phenomena. I think that his denotation and vividness conditions, and even his of-ness condition, are not right. These drawbacks seem to me far less important than his valuable initiative in exploring epistemic intuitions about cognitive states that go beyond conceptualization or description.

In "Belief *De Re*" (1977), I criticized some theses of Kaplan's paper and noted some ways in which its linguistic focus blurred a clear view of the epistemic basis for the distinction.⁶⁰ I centered the account more on epistemic considerations. That paper is the basis for the reflections in

warrant for certain applications of spatial concepts is not apriori. This is, however, a complex and difficult issue.

57. Bertrand Russell, "Knowledge by Acquaintance and Knowledge by Description," in *The Problems of Philosophy* (1912; Oxford: Oxford University Press, 1982).

58. W. V. Quine, "Quantifiers and Propositional Attitudes," in *Ways of Paradox* (New York: Random House, 1966).

59. David Kaplan, "Quantifying In," in *Words and Objections: Essays on the Work of W. V. Quine*, ed. D. Davidson and J. Hintikka (Dordrecht: D. Reidel, 1969).

60. Cf. my "Belief *De Re*." The point about separating linguistic phenomena from facts about *de re* attitudes is made in that article. The criticisms of Kaplan's use of the notions of denotation and vividness are also laid out in that article. *Denotation* is not explicitly contextual in his early work. I argued that the context-dependence of applications is key to paradigmatic *de re* cases. *Vividness* seems to me clearly unnecessary to *de re* states or attitudes, even for *autonomous* thought. Vividness is, in my view, an empiricist red herring. Many *de re* attitudes are not vivid, and many vivid attitudes are not *de re*. Vividness does not constitutively bear on the character of a representation's relation in being of (*de*) a subject matter (*re*). There are straightforward counter-examples to vividness as a necessary condition. One can form a perceptual belief of an object (or other particular) and not register or remember enough of its features to have a vivid representation. Such beliefs

the present work. But in my 1977 paper I, too, leaned excessively on linguistic phenomena.⁶¹

Although nearly everyone, from Russell onward, took perception as the paradigm of *de re* phenomena, most conceived the distinction mainly in linguistic terms. Given that linguistic attribution of attitudes is subject to pragmatic pressures other than specifying the types of attitudes being attributed, there is no simple correlation between types of attribution (showing logical features like those that interested Quine) and types of state. The gradual realization of this fact led to a malaise. Some wondered whether there is any *de re-de dicto* distinction at all.

The initial lesson here is easy and old. Look not to ordinary language for immediate or final insight into the nature of things. Ordinary language is busy with too much else to provide unstinting service to philosophy or science, insofar as they are concerned with something beyond language itself. This is not to say that the nature of things is always esoteric or surprising, or that language does not yield insight. It is just to say that linguistic attribution of cognition and cognition itself are really quite different matters, with only complex relations between them.

Even with the lesson assimilated, one can find it hard to decide what should be understood by the *de re-de dicto* distinction. There are many distinctions in the area. Some grade off into vagueness after a few clear cases. Some clamor against one another to be attached to the famous terms. I believe that it remains a fruitful enterprise to seek a distinction connected with the terms that is conceptually rich, but is clear enough to serve philosophy, and perhaps even science.

I began with two of Russell's ideas. One is his idea of a representational state that is not purely descriptive. The other is his idea of perception as paradigm. Combining the two ideas, I began by reflecting on not-purely-descriptive aspects of perception. Perception does involve more than the analogs of descriptions. It involves context-bound singular elements guided by but not replaceable by nonschematic, semantically general, ability general attributive representations.

When we visually represent a scene, the visual system contributes ability general representations that attribute kinds, properties, and relations. These representations cannot be all there is to perceptual representation. If

are common, and even basic to action. They need not even be conscious, or driven by unconscious attention. Clearly one can have a *de re* belief of an object (or other particular) in such cases. The epistemic relation is very direct; it is only partly conceptualized; and it is context-dependent. Moreover, vividness is irrelevant to *de se* or egocentric indexes, which seem clearly to be subspecies of *de re* reference. (The distinction between *se* and *re* is not ontological. *Se*'s are *re*'s. The distinction lies in the mode of presentation.) In effect, I criticize the causal *of-ness* condition—Kaplan's third condition on *de re* thought—in section V of this essay. The basic idea of reference that is backed by an immediate nonconceptual representational and epistemic capacity, which grounds my present positive account, is initially developed in "Belief *De Re*."

61. Cf. my "Postscript to 'Belief *De Re*,' " in my *Foundations of Mind*.

a perceptually indiscernible scene were to be somewhere else in the universe, one would perceive the scene that causes one's perception, not the duplicate scene. Intuitive and scientific considerations rule out attributing to the perceptual system representations like whatever causes this representation (cf. note 48). The perceptual system cannot itself discern the difference between the two scenes by means of its general representational abilities. Since representations function partly to mark ability, the general representational abilities should be type-identified or marked by semantically general representations, which apply to both scenes. The perceptual system functions to represent entities relevant to the individual's functions. In the case of hypothetical duplicates, the individuals see and perceptually represent particulars that cause their perceptions in the context. So a context-bound, semantically singular element is needed to account for the perception's (fallibly) representing the particulars that cause it.⁶²

I took perception to be a paradigm *de re* state. An initial hypothesis arose from reflection on this paradigm. The *de re* nature of the states constitutively depends on their being partly type-identified by context-bound singular representations (applications) that do not rely purely on nonschematic, semantically general or ability general attributives for their representational success.

Extending this initial paradigm to perceptual memory, to perceptual belief, and to belief based on perceptual memory would raise many interesting issues of detail. I think, however, that the basic form of the extension is not hard to see. All such states have in their representational content a singular representational element, inherited from perception, that marks a fallible representational ability that is context-bound, not ability general. All such *de re* states and attitudes involve representational abilities that are singular and context-bound. This is the analog of the more linguistically oriented dictum: Showing beats telling.

Before proceeding, I want to flag an issue that I will not pursue here in depth. Strictly speaking, to be *de re*, a state or attitude must *succeed* in referring to a *re*. Seeing requires referential success and is paradigmatically *de re*. Is seeing a psychological state or attitude?⁶³ Ordinary language is liberal with state talk. Perhaps it is a state. I doubt, however, that it is a fundamental explanatory kind, as opposed to a kind to be explained, in psychology. Even if it is, there are psychological kinds that include both seeings and perceptual, even referential, illusions. Such kinds figure in explaining seeing.⁶⁴

62. The main idea of this argument derives from Peter Strawson's brilliant duplication argument, *Individuals* (1959; Garden City, N.Y.: Doubleday, 1963), chap. 1.

63. Cf. Timothy Williamson, "Is Knowing a State of Mind?" *Mind* 104 (1995): 533–565. It would be a mistake to construe the view that I develop in what follows, and elsewhere, as regarding seeing as *analyzable* into visual representation and causation.

64. For an extensive discussion of this issue, see my "Disjunctivism and Perceptual Psychology."

Explanations in psychology fix on perceptual states that in normal conditions constitute seeing. They are motivated by the phenomenon of seeing. They begin by explaining the successes. Anti-individualism takes seeing to be the phenomenon that underlies the determination of ability general visual representations. But the methods and explanations of psychology often count states the same in conditions when the individual and perceptual system are fooled. It is central to the methodology of the science of vision that this be so. There are solid general empirical reasons for this methodology that I shall not go into in detail here.⁶⁵ Briefly, psychological kinds involve the processing of perceptual representations according to certain principles that come into play given stimulation of the retina. These principles hold regardless of whether the stimulation derives from a *re* in the normal way that makes successful perception possible. In cases where the representations arise from contextually abnormal distal conditions, the psychological processing may remain the same. Perceptual states are individuated in psychology to allow the same kind of state (at one level of kind-individuation) to be the same whether it is veridical or illusional. Explanatory successes in the psychology of vision have been united in following this methodology.

The difference between successful perceptual reference (or seeing) and perceptual referential illusion can be serendipitous. The difference can turn on the whim of the experimental psychologist. Fundamental psychological explanation abstracts from such vicissitudes. Even if seeing *does* turn out to be a psychological kind in this narrow sense, it is clear that there are explanatorily relevant psychological kinds that are not factive, as seeing and knowing are.

I am interested in the broader array of psychological states that help explain seeing, even though not all are successfully “of” a *re*. Seeing and other strictly *de re* phenomena are explicitly relational kinds.⁶⁶ They are real. They are in some ways fundamental. They motivate the explanatory kinds that psychological explanation actually uses. These kinds have the same form as strictly *de re* phenomena, but they do not require referential success. They constitute an important psychological kind. When I write of *de re* states or attitudes, I mean that they are proleptically *de re*: They are states and attitudes of a sort that when successfully referential are *de re*.

65. Cf. “Disjunctivism and Perceptual Psychology.”

66. For earlier statements of this view, see “Belief *De Re*.” Note that knowings are not the only *de re* propositional phenomena under the strict usage. One can have a *de re* belief that is successfully referential and meets all other conditions on being *de re*, which nevertheless fails to count as knowledge. Suppose that one is looking directly at an object and that one forms a true belief about it. Suppose that one has good reason to doubt that there is really an object there. For example, suppose that one is in a psychological experiment where one has good reason to believe that one has been fooled frequently. Suppose that one ignores this good reason. Then one lacks knowledge. But one has a *de re* belief of the object. I have not been able to think of any purely visual states that are strictly *de re*, and therefore successfully referential, which are not also seeings.

Let us return to the issue of what if any states and attitudes to count as *de re*—beyond perceptions, perceptual beliefs (and other perceptually informed attitudes), and memories of all these. “*De re*” is a term of art. One could stop here. I think, however, that there are further cases that belong among mental states or attitudes that can reasonably be counted *de re*. We should avoid the empiricist presumption that the only sort of not-purely-descriptive representational or epistemic relation that we have to a *re* is through perception. Avoiding this presumption leads to a range of interesting phenomena that have some of the “directness” of the perceptual, but that are not empirically based and not dependent on causation in the same way that perception is.

All cases of *de re* states and attitudes so far discussed have featured causation by the referent. I think that there are *de re* states and attitudes that do not have this feature.⁶⁷ I begin simply by collecting some examples. Collection will continue to be guided by Russell’s idea of reference to an object via not-purely-descriptive means. There are at least four types of cases.

One type involves uses of simple indexicals in thought. My occurrent thought I am thinking seems clearly *de re* with respect to me. The referent of I is not fixed by some event in me causing the occurrence of I. It is fixed by my authoring the thought. The referential and epistemic access to myself in such a case is not essentially empirical. I can know empirically that I am thinking. But referential and epistemic access to myself need not rely on empirical means. I may have already identified myself through my awareness of the thinking. If the empirical information I had about myself were mistaken, I would still succeed in representing myself with I. Access goes through a framework role for I and through intellectual access to my occurrent thought. Neither the framework role nor the awareness of my thinking is reducible to empirical or other causal paradigms of reference.⁶⁸

I think that similar points can be made for normal uses of now, and some occurrences of here, in thought. A thought it is now raining is normally *de re* with respect to the present moment. The referent is fixed neither by some context-free description nor by the present moment’s causing the occurrence of the indexical. It need not be fixed

67. What should we say about uses of names of individuals that one has never perceived—“Aristotle” or “Ninevah”? Kripke and Donnellan showed that such names can refer even though their user lacks descriptions sufficient to fix their referents. There is a directness and noninferentiality to the understanding of such names that makes it kin to the perceptual paradigm, despite the poverty of information and distance in history. I believe that thoughts making use of such names can be considered *de re*. They have a special status, however. They are *de re* only nonautonomously—only through reliance on others.

68. Cf. my “Reason and the First Person,” in *Knowing Our Own Minds: Essays on Self-Knowledge*, ed. C. Wright, B. C. Smith, and C. MacDonald (Oxford: Clarendon Press, 1998), and “Memory and Persons.”

through perception of other things. It is fixed by context-bound application of the schematic concept now.

Reference through such indexicals is certainly not purely descriptive. Context-bound singular application is necessary. Our epistemic access through indexicals to ourselves, to the present time, and often to the present place is not purely a matter of perception. There need be no separate faculty of apprehension of the referents. The epistemic access is associated with the mastery of certain frameworks and systems of coordination—including general egocentrically oriented systems of action, and general temporal and spatial abilities. These frameworks mark, at their *de se*, spatial, and temporal anchor points, immediately applicable cognitive and practical abilities.

The range of “indexical” referential phenomena is wider and more primitive than the cases just mentioned may suggest. Many animals that lack propositional attitudes have perceptual systems and activities geared to their perceptions. Egocentric indexes that are relevant to action (fleeing, eating, mating perceived objects) are built into the framework of all perception and action. Framework-origins of temporal and spatial perceptual frameworks are associated with the egocentric indexes. These are primitive analogs of the conceptual indexicals, I, now, and here. These indexes represent their referents not through causal relations but through context-dependent orientation of the frameworks that they anchor in perception and primitive agency.

These markers’ referential success does not depend on a present perceptual or other causal relation to the “referents” that they index. The referential link is established in having and using competencies constitutive of a representational perspective. All *de re* representation in states and attitudes, even in perception, hence *all* representation, presupposes that these direct, noninferential, nondescriptive links are in place.⁶⁹

A second group of cases that are plausibly *de re* but where causation is not necessary for reference comprises certain types of self-knowledge of one’s mental states and events.

Some self-knowledge is empirical and causally based. One can know one’s mind from the outside by observing oneself. Even some authoritative self-knowledge has a causal base. My belief that I have a memory as of hearing Rubinstein play Chopin’s Third Scherzo might be based on the belief’s being caused by the memory. My belief that I am in pain might be based on the pain’s causing the belief.

Some of our self-knowledge, however, is neither warranted through perception nor dependent for successful reference on being caused by the mental events that are known. In the *cogito* thought I am hereby thinking that music is valuable there are noncausal representational relations to the author of the thought and the present time. These are *de re* indexical references of the sort already discussed. The *cogito* thought also contains

69. Cf. “Memory and Persons,” especially section V.

de re reference to the event of thinking the thought. The representational relation to the thought event is not caused by the thought event that it is about. The reference depends on the mental activity and on the form of the thought, not on a causal relation between *re* and representation. The knowledge is intellectual, not perceptual or causal. Understanding the thought that one is thinking suffices for knowing that it is occurring.

A third type of case is closely related to the second. It seems to me that one can have not-purely-descriptive referential attitudes toward actions that one intends and that one is about to carry out.⁷⁰ I can think of this (coming) raising of my arm just before I raise it. I believe that successful reference need not rely on a description like “the action that I am about to perform.” It can rely on the competence routines and power that will issue in the act. A pastor might in marrying a couple say, “Let no man put this marital union asunder.” At the time of the application of “this marital union” there may not yet be a marriage. Intentional control over the future can yield not-purely-descriptive, noninferential representational relations to an object or event. I think it reasonable to count such relations *de re*.

How much control is necessary? How far into the future can such attitudes reach? Perhaps answers will never be sharp. I think, however, that noninferential cognitive relations to future entities that are under reliable representational and practical control have an epistemic directness and an independence of context-free conceptualization that make them hard to exclude from the representational and epistemic phenomena that Russell opposed to knowledge by description.

A fourth candidate type of noncausal *de re* attitudes comprises certain cognitive relations to abstract entities. Russell counted grasp of universals an acquaintance relation. I believe that this position resulted from his characteristic conflation of understanding with referential relations to objects. In predicating a concept of an object in the thought that man is a great pianist, we think the concept is a great pianist as part of thinking the thought. Thinking the concept is not a representational relation to the concept. The thought is not about (*de*) the concept. The relation should not be counted *de re*.

There are, however, cases where comprehension and reference are inseparable. Attributions of thought normally contain specification of the thought in a canonical way that requires thinking the thought content as one attributes it. When I think I (or you) believe that not all people are great pianists, I must think the representational thought content that not all people are great pianists in the course of attributing it. I also canonically *name* or *designate* the representational thought content via a

70. This is perhaps a distant analog of intellectual intuition, attributed by the medievals to God—an ability to intuit objects and thereby create them. Cf. note 3. We do not create by intuiting. But the intuition may be guided by the creation—the forward-looking causal power. These cases are interestingly discussed by G. E. M. Anscombe, *Intention* (Oxford: Blackwell, 1957), though not by reference to the notion of *de re* states.

singular term, the analog in thought of a that-clause.⁷¹ My relation to the referent is not purely descriptive. It is true that the canonical specification is ability general and conceptual. But the specification is backed by *comprehension* of the referent. Comprehension is at least as direct and noninferential, psychologically and epistemically, as perceptual relations. Comprehending a representational content is exercising an ability that is constitutively associated with inference. But it is not itself inferential or descriptive. I think that comprehension is a direct intellectual capacity that *when constitutively combined with reference* can make *de re* reference possible, when reference is carried out in this canonical way.⁷² I think it clear that representational contents to which we bear these *de re* relations do not cause the reference. The *de re* representation is not empirical. It is intellectual, though *some* of the relevant *de re* thoughts are warranted empirically.

Canonical specification of simple natural numbers through numerals is also arguably *de re*. (I assume a realist attitude toward the numbers. Anti-realists may form whatever conclusions they will.) We do not perceive the numbers. They do not cause our thought about them. Numerals in a canonical system contrast with nonmathematical descriptions ("the number of cats my sister has") or computationally difficult mathematical formulas. They enable one to relate any complex name by simple mechanical means to the simplest numerals. The *basic* elements of the system are repeated in combinations to form larger groups. These basic elements are like the indexical origins (or *de se* origins) of spatial or temporal frameworks. They are the starting points that we use, together with general operations, to specify other "points" (on the analogy to spatial locations or times) in the numerical system. Our ability to specify 1, 2, 3, 4, 5... (certain among the smallest natural numbers) through simple

71. A special feature of these (e.g., that-clause) canonical content-names is that mastering them requires mastery of the named or referred-to contents themselves. So there is, in a certain way, an even more intimate relation between this sort of canonical name and its named contents than there is between a canonical number name like "2" and the number. Here one literally must understand the denotation (the customary content or sense) before grasping the content of the name or individual concept that canonically names it. Grasp of the denotation or referent precedes grasp of the content that represents it. For further discussion, see my "Postscript to 'Frege and the Hierarchy,'" in my *Truth, Thought, Reason: Essays on Frege* (Oxford: Clarendon Press, 2005).

72. A more empirical case involves thought about color types represented in perception or memory by a color-sighted person. The color type does not cause the thought. It has no causal power. Only instances of the color type can cause anything. One might think of the color type *without remembering any instance*. There does remain some causal relation back to instances in the learning history. But it is implausible to think that any given instance caused this occurrence of thought. Moreover, one might imagine a color shade even though one never saw an instance. One might imagine the color while one is thinking of it. The image itself does not have the color. Here is *de re* thought without direct causal connection. I think that one could think *de re* of the color without imagining it at all. The *power* to imagine it seems arguably sufficient. I am indebted to Mark Johnston for the idea of this note, though not the details.

words or noncomplex numerals, has an epistemic primitiveness that is relevant in determining what should count as *de re*.

There is evidence that some abilities with small numbers are universal among humans, despite differences in symbolic systems. For example, humans can determine correlations between images or perceptions of groups of objects and these numbers very quickly, without counting or calculating. This ability is widely studied in cognitive psychology under the rubric *subitizing*.⁷³ In fact, perceptual subitizing is common throughout the animal kingdom. Of course, the perceptual system computes, but these computations are modular. The *individual's* noninferential recognition of the number of a small group of items is approximately as immediate as any perceptual representation. Subitizing is not perception of abstract objects, the numbers. But in individuals who have an understanding of a numerical system, the primitive subitizing capacities join with conceptual abilities to support noninferential, noncomputational numerical assignments in thought to small groupings. These assignments are associated with noninferential conceptual ability to use canonical specifications of these numbers as *bases* for computations (that is, without representing these numbers as the products of computations). So the representation 2 is primitive—in contrast to compounds like the successor of 1 or 12.

These noninferential representational and applicational abilities are the basic elements in a great deal of mathematical knowledge. Resolution of computations into basic psychological and epistemic elements offers a ground for understanding effectiveness (or effective calculability), mathematical proof, and so on. Thus certain small natural numbers, though certainly specified conceptually, can be naturally associated both with immediate conceptually aided perception of groups as having those cardinalities and with immediate (noncomputational) representation of numbers in pure, nonapplied arithmetic. I conjecture that it is reasonable to count representation of mathematical objects that is backed by such noninferential abilities of application and understanding as *de re* with respect to those objects.⁷⁴

73. A. Klein and P. Starkey, "The Origins and Development of Numerical Cognition: A Comparative Analysis," in *Cognitive Processes in Mathematics*, ed. J. Sloboda and D. Rogers (Oxford: Oxford University Press, 1987); Karen Wynn, "Psychological Foundations of Number: Numerical Competence in Human Infants," *Trends in Cognitive Sciences* 2 (1998): 296–303; J. Whalen, C. R. Gallistel, and R. Gelman, "Nonverbal Counting in Humans," *Psychological Science* 10 (1999): 130–137; M. C. Uller, S. Carey, G. Huntley-Fenner, and L. Klatt, "What Representations Might Underlie Infant Numerical Knowledge?" *Cognitive Development* 14 (1999): 1–36. Subitizing occurs in lower animals that lack propositional attitudes. I believe that subitizing does not itself make reference to numbers. It certainly occurs phylogenetically earlier than even the most primitive mastery of the arithmetical system of numbers. But I believe that it is a source of immediacy in arithmetical cognition.

74. Kaplan, in "Quantifying In," section VIII, discusses canonical names, calling them "standard names." He counts quotation names and numerals as examples. He centers on

What can be immediately, noninferentially surveyed may vary with expertise and ability. What impresses me is that there is evidence that there are relatively sharp and universal boundaries between those number specifications that can and those that cannot be applied (or used in pure arithmetic) by ordinary people without counting.⁷⁵

I have taken as key to the *de re–de dicto* distinction Russell's idea that *de re* states and attitudes involve a capacity for referring to entities that is essentially nondescriptive, noninferential, and epistemically immediate. Perception, perceptual belief, and perceptual memory provide a start toward understanding *de re* states and attitudes. I maintained that resting there would be to accept a narrow empiricist conception of our basic cognitive and representational capacities. I believe that we have *de re* representation *through understanding*, not just perception.

I outlined four capacities for referential representation that seem to go beyond the perceptual paradigm. All involve not-purely-descriptive representations of objects. All go beyond use of ability general, purely descriptive representations. All are backed by epistemic capacities that are noninferential, immediate, and nondiscursive.

Representation with certain indexicals and *de se* markers seems to be associated with a nondescriptive setting of the origins of representational frameworks.

Representation of mental states and events in reflective self-attributions is associated with noninferential epistemic relation that is context-bound, singular, and not purely descriptive.

their modal properties and does not elaborate their epistemic properties, though he compares his vivid names to standard names in section XI. I want to emphasize that I do not hold that all uses of canonical names are associated with *de re* attitudes. Uses of large numerals commonly are not. The key issue concerns the immediacy, the noncomputational and noninferential character, of the individual's representational and epistemic capacity that backs use of the name. Thus I think that, normally, only uses of very small numerals yield *de re* attitudes. I believe that uses of canonical names for noncomplex representational contents and for relatively simple combinations of them are candidates for yielding *de re* attitudes. For example, representational contents of that-clauses containing obliquely occurring expressions that can be comprehended without exercise of nonmodular computation can involve *de re* representation. Here again see my "Postscript to 'Frege and the Hierarchy.'" For a fuller discussion of my particular view of *de re* thought about small natural numbers, see my "Postscript to 'Belief De Re.'"

75. The third and fourth types of nonempirical *de re* cases raise interesting questions about reference to the future. Many references to objects in the future do not support *de re* attitudes. Reference through complete definite descriptions is, of course, an example. Equally, names introduced in terms of context-free definite descriptions ("Newman 1 and 1/2"—introduced as referring to the first person born in 2050) are examples. The same can be said about indexically infected descriptions like "the 754,573,211,467th day after today." Such a description can support an attitude that is *de re* with respect to today, but not with respect to the 754,573,211,467th day after today. The reason is, again, not absence of a causal relation. It is that the relation to that future day is, apart from the indexical anchoring in today, *entirely* dependent on ability general representation.

Representation of intended acts or objects over which one has control seems associated with a noninferential ability to know and represent them by nondescriptive context-bound singular means.

Canonical representations of understood representational contents and certain canonical representations of simple natural numbers are candidates for *de re* status. Unlike the representations in the other cases, the relevant representations of these abstract entities are *ability general* though semantically singular. The representations are fully conceptual.⁷⁶ Such representation occurs within canonical systems of designation that do not themselves rely on context-bound forms of reference. In these respects, reference here is significantly different from other *de re* reference. Still, the canonical system of representation is, at its bases or origins, intuitively nondescriptive. The basic canonical representations are also partly backed by noninferential, noncomputational modes of reference and understanding.

I accept Frege's point that we do not know the numbers through perceptual-like apprehension of them. We know them only through understanding arithmetical propositions. Here, reference derives from propositional abilities, not from a subpropositional ability like perception. Still, comprehending the thoughts that canonically specify the smallest natural numbers through numerals is essentially linked to a noninferential representational ability—the conceptualized successor of

What are we to say about attitudes using applications of the indexical tomorrow? Kaplan denies attitudes *de re* status with respect to positions in which tomorrow occurs. He appears to base this denial on our lack of causal relation to future days. Cf. David Kaplan, "De Re Belief," (2003) in *Presidential Addresses of the American Philosophical Association*, ed. Richard T. Hull (forthcoming). This may signal a different conception of *de re*. From my perspective, a denial of *de re* status based on absence of a causal relation to the *re* would not be a good reason. We have seen counterexamples to this principle from other quarters. The indexical tomorrow depends for its referential workings on a relation to today. It refers to the day after today. In this respect, it is like the description of a future day that uses a huge number. On the other hand, often we can have virtually as direct an epistemic and representational relation to tomorrow as we do to today—if we are thoroughly centered on our plans for tomorrow, for example.

I believe that indexicals like tomorrow can yield states and attitudes that are *de re* with respect to future times. Their being single words suggests that no inference need be made in their application. Their being single words is not decisive, of course. One could coin a one-word indexical for the 754,573,211,467th day after today. Because most of us cannot parse or apply the number noninferentially, such an indexical could not be used to think *de re* thoughts with respect to the relevant day. What enables tomorrow sometimes to effect *de re* reference is that the day is often cognitively and practically at hand. This is partly because the day bears a numerical relation to today that itself does not require inference or counting for its application. It is partly because we have power over our acts in the immediate future. The issues again invite further reflection—on another day.

76. This point constitutes a departure from one of the lines on *de re* attitudes that I took in "Belief *De Re*." For a criticism of that line and further motivation for the line taken here, see "Postscript to 'Belief *De Re*.'"

subitizing. This is recognition and application of numbers without calculation or description. It is recognition through singular understanding.

A similar point applies to our knowledge of thought contents. We do not know them through perceptual apprehension. We know them through reflexive, meta-representational specification of what we discursively understand. The basic non-meta-representational, discursive, competence understanding is a combination of applicational ability, attributional and recognitional ability, and inference. But meta-representational understanding of a content through canonical names is not description or inference. The ability to canonically name representational contents that we have a competence-understanding of is a nondescriptive, non-inferential, nonattributional ability. And the ability to think, with understanding, about contents thus canonically named need not employ inference. Thoughts that exercise that ability are, I think, *de re*. They constitute another type of singular intellectual understanding.

The fifth thesis is:

A mental state or attitude is autonomously (and proleptically) de re with respect to a representational position in its representational content if and only if the representational position contains a representational content that represents (purports to refer) nondescriptively and is backed by an epistemic competence to make noninferential, immediate, nondiscursive attributions to the re.

In sufficiently mature thinkers, exercise of this competence often constitutes knowledge. It can, however, reside in primitive, subpropositional perception or action and in framework-setting *de se* markers in perception or action sets.

I have acknowledged many issues that challenge further reflection. I hope to have indicated that understanding *de re* phenomena is a project not only in the theory of reference, let alone belief-attribution. It is a project that probes fundamental epistemic and representational capacities that underlie what it is to have a mind.

INDEX

- ability generality. *See* generality, ability
- ability particularity, 260–266, 265ff, 284–285
- accompaniment, 272–275, 281, 296
- accuracy, 253, 289. *See also*
 - veridicality, perceptual
- acquaintance, 246–247, 251, 305, 311
- action-awareness, 178–182
- “Afterthoughts,” 187, 189–190, 199
- algebraic approach, 109, 114, 117, 119–121, 130–131
- Allen, Woody, 31
- anaphors, 200–201
- Anderson, C. Anders (Tony), 8
- Anscombe, Gertrude E. M., 311fn
- anti-individualism, 288, 297, 308
- antinomy (puzzle) of the variable, 109–112, 114, 116–117, 122, 125, 130
- application, 250, 259–261, 263, 286, 307, 310
 - attributive, 255, 304
 - in thought, 256, 273–274, 285–286, 297
 - perceptual, 256, 271, 273–274, 283, 286, 301–302
 - plural, 255–256
 - singular, 250, 253–256, 259–261, 268–269, 287, 301
- apriori knowledge. *See* apriori warrant
- apriori warrant, 247, 283, 297–304
- Aristotle, 261fn, 278fn
- armchair reflection, 303
- assertion, 152
- attitude attribution, 305ff
- attribute, 254
- attribution, 250, 252–288, 278, 288–297, 290, 297–304
 - purported, 250, 253–254
- attributive
 - illusion, 254
 - in thought (predicative concept), 247, 253
 - perceptual, 247, 253, 259–262, 265
- Austin, John L., 4
- automorphism, 75
- autonomous thought, 274–275
- Ayer, Alfred J., 17, 30

- Barcan Marcus, Ruth, 32
- Bach, Kent, 153
- background assumption, 176–177
- Barker, Stephen, 154
- belief(s), 190, 192–195, 198, 233–245
 - notional, 35, 38, 235–240, 242
 - relational, 35, 38, 235–236, 238–240, 242
- belief and modal logic, 13
- Belnap, Nuel, 19, 138–139, 144
- Bernays-Morse set theory. *See* Kelly
 - Morse set theory
- body, representation of, 257, 280–281, 292, 299–304
- Boghossian, Paul, 140–141
- boolean functions, 63
- Boone, William W., 16
- Brandom, Robert, 139–140, 143
- Burge, Tyler, 26, 53–55, 57, 160, 162, 173, 184, 186, 241

- canonical concept of a concept.
 - See* concept of a concept, canonical
- Canonical Form Assumption, 209–210
- canonical name, 311–316
- Cantor’s diagonal argument, 86
- Carnapian propositions, 13, 15, 85–92
- Carnapian semantics, 45, 49

- Carnap Rudolf, 3–4, 6–7, 12, 16–7, 27, 29–30, 88, 192. *See also*
 Carnapian propositions
- Carston, Robin, 153
- causation, 277–278, 284, 286, 290, 293–295, 297, 306, 309–315
- character, 5, 188, 196, 200
 and content, 22
- Church, Alonzo, 3–4, 7–8, 16–7, 26, 30, 32, 88
 and Kaplan, 22–23
See also Frege-Church approach to intensional logic; Frege-Church-Gödel slingshot
- cognitive modality, 233–234
- Cohen, L. Jonathan, 89–90
- color, representation of, 279–280, 288, 295, 300–301, 303, 312
- comprehension, 311–312, 315. *See also* understanding
- computability, 312–315
- concept(s) (conceptual representation), 137, 139–141, 143, 154–155, 160–178, 180–185, 247–248, 253, 259, 261–263, 273–274, 278, 287, 298, 306. *See also* individual concept
- concept of a concept, canonical, 164–173, 175–177, 180–183
- concepts of concepts, 162–163, 165–166, 168, 171–172, 182
- conscious states (and events), 169, 182
- conservative extension, 138–140, 145–147
- containment, 253, 263–264, 270–274, 278
- content, 188–189, 192–195, 234, 237–238, 240–244
 intentional, 160, 177–180, 182
 locutionary (*see* locutionary content)
- context(s), 187–188, 193, 200–201
 indirect (*see* indirect context(s))
 nonextensional, 38
 of use, 4–5
See also opaque contexts
- context-boundedness. *See* representation, context-bound
- conventional implicature, 149–155
- conversational implicature, 150, 152, 155
- coordination (coordinated, coordinating), 128–132
- coreference
 accidental, 131–132
 strict, 131–132
- Craig's Interpolation Lemma, 19
- Davidson, Donald, 160–161, 163, 287fn
- Davies, Martin, 18
- Davis, Angela, 31
- definition, contextual, 37, 38fn
- demonstration, 188–190, 192
- demonstrative(s), 4–5, 187–190, 192–193, 196, 198, 200–201, 250, 255–256, 259–261, 263–264, 267–268, 270, 274–275, 278–279, 282–283, 298–302
- "Demonstratives," 4, 187, 189, 196, 200, 213fn 7, 215, 290fn, 301fn
- de dicto* psychological description, 221–223
- denote (denoted, denotation), 237, 239, 241–242, 305, 312
- "De Re Belief," 315fn
- de re/de dicto*, 21, 32
de dicto, 35–36, 38, 38fn, 172, 176
de re, 35–36, 38, 38fn, 176–8, 246–247, 297–298, 305–316, 246–316 *passim*
de re belief 233–234, 238–244
de se marker, 256, 263–264, 269–271, 306, 310, 312, 314, 316
 (*see also* egocentric index)
- pseudo *de re* belief, 244
See also de dicto psychological description
- Descartes, René, 246fn, 262fn
- description, 31, 33–34
- determiners
 cardinal, 67
 co-cardinal, 68
 co-intersective, 68
 generalized existential, 66
 generalized universal, 68
 intersective, 66
 k-place, 71
 non-logical, 75
 permutation invariant, 75
 proportionality, 69, 70
 unreducible, 74

- determinist theory, 20
- Deutsch, Harry, 26
- diagonal argument, 101–105
- directing intention(s), 189–190, 194, 196, 198–201
- Directing Intentions Theory (of demonstratives), 189–190
- direct reference, 6, 14, 32, 246, 249–252
- “Dthat,” 7, 189
- Donnellan, Keith, 5, 7–8, 26, 32, 39*fn*, 41, 233,
- Dummett, Michael, 139–142, 144–145, 147–149, 154, 160, 162–163, 172–173, 175, 183, 289, 289*fn*, 290, 309*fn*
- duplicate, 284–285, 306–307
- Eaker, Erin, 8
- Eberle, Rolf, 19
- effective computability. *See* computability
- egocentric index, 256, 263–264, 269–271, 306, 310, 312, 314, 316
- empirical significance, 30
- empirically committal representation, 289–290
- empiricism (empiricist), 247, 297–304, 305, 309–316
- Epimenides paradox, 87–92
- epistemic norm, 249
- epistemic relation (direct, non-inferential relation), 246, 305–306, 309–316
- epistemology, 246, 251, 288, 296, 303, 305
- essentialism, 6, 16–17
- Evans, Gareth, 9, 287*fn*, 291*fn*
- explanation, 19
- exportation of a concept. *See* concept, exportation
- expression
 - closed, 114–115, 117
 - open, 114–117, 127, 130
- extensional position, 59
- external, 166, 168, 178–181
- extrinsic, 123
- fallibility, 250–253, 277–278, 284–288, 299
- free repeatability, 259–261, 265, 276–279, 283, 285–288, 293
- Fine, Kit, 14*fn*
- first-order logic
 - language, 126
 - semantics, 109, 132
 - symbolism, 120–121, 125
- Fitch, Frederic, 90, 90*fn*
- Fodor, Jerry (Fodorian), 167–169
- Frege, Gottlob (Fregean), 4–7, 11, 14, 16, 31, 36, 38–39, 41, 45, 47–49, 53, 56, 62, 63, 88, 109–110, 118–119, 131–132, 144, 148–149, 161, 169, 173, 183, 202, 233, 238–239, 241–243, 245, 249*fn*, 263*fn*, 312*fn*, 314*fn*, 315
 - Fregean sense, 33
 - Fregean thought, 36, 38
 - semantics for belief sentences, 210
 - theory of *Sinn* and *Bedeutung*, 202–204, 206
- Frege-Church approach to intensional logic, 11–12, 14–15, 21
- Frege-Church-Gödel slingshot, 11
- Frege’s hierarchy of senses, 160, 162–163, 165, 167, 169–170, 172–176, 183–184
- Frege’s puzzle, 110, 131, 203–204
- Fregean Theory (of demonstratives), 187–190
- future referents, 311, 314–315
- generality, 251–254, 256–258, 276, 281–282, 286–291, 293–294, 296, 300, 307
 - ability, 259–266, 268–272, 276–286, 308, 312, 314–315
- quantificational, 258
- schematic, 261, 263–270, 278, 283–284
- semantic, 246, 258, 261–272, 275–278, 288, 290, 307
- syntactic, 258, 262–269, 278
- gesture(s), 189–190, 193–194
- God, 251, 311
- Gödel, Kurt, 90
- Grice, H. Paul, 149–150, 152
- guidance, 252, 254, 256–258, 264, 269–275, 279–284, 287–297, 300–303

- Haecceitism, and anti-Haecceitism, 21
 Hare, Richard M., 151
 harmony (proof-theoretic), 139–140, 145
 Hempel, Carl, 19
 hierarchy of senses. *See* sense(s), hierarchy of
 Hintikka, Jaakko, 16
 Horn, Larry, 144
 Hornsby, Jennifer, 151, 154

 identity sentences
 epistemic status of, 203
 forms of, 202–205
 illusion. *See* referential illusion;
 attributive illusion
 implicature. *See* conventional
 implicature; conversational
 implicature
 independent semantic role, 116–117, 124–125, 127
 index, 6
 indexical(s) (indexically), 187, 189, 194, 196, 256, 258, 261, 263–264, 267, 270, 285, 298, 309–312, 314–315
 indexicalism, 298
 Indexical Theory (of demonstratives), 187–188
 indication, 254, 268
 indirect context(s) (quotation), 46, 48–49, 53–54, 57–58
 individual concept(s), 14–5, 21, 265, 268, 275, 312
 inference, 264–5, 274, 288, 297–298, 309–316
 and rules, 137–147
 inferentialism, 137–140, 142–143, 147–149
 information delimitation, 153
 instantial approach, 109, 114, 117, 119, 130
 intentional content. *See* content, intentional
 intentionality, 234, 239–240, 242*fn*, 243–244
 intensional Liar paradox, 93, 98–99, 101–105
 intensional logic, 94–6

 intention(s), 189–195, 197–199, 201.
 See also directing intension(s)
 internal (internalist), 178–180, 234, 238–243
 interpreted logical forms, 212
 intrinsic, 123–126, 130–132

 Johnston, Mark, 312*fn*
 Jubien, Michael, 18

 Kalish, Donald, 26, 27, 29
 Kamp, Hans, 5, 8, 21
 Kant, Immanuel, 246, 278*fn*, 304
 Kaplan, David, 3–10, 11–23, 25–33, 35–41, 45, 59, 80–81, 85–92, 93, 95–8, 100–101, 106, 132, 142, 153, 156, 159–160, 171, 202, 203*fn* 4 and 7, 210, 249*fn*, 290*fn*, 301*fn*, 305, 305*fn*, 313*fn*, 315*fn*
 Kaplan's paradox, 93, 96–98, 101–105
 Karttunen, Larry, 151
 Kelly-Morse set theory, 20
 kind, 248–249, 251, 254, 257–259, 280–284, 288, 291, 293–295
 knower, paradox of, 18–19
 knowledge. *See* apriori warrant
 knowledge by acquaintance, 39
 knowledge how, 142–143
 Korzybski, Alfred, 87
 Kripke, Saul A., 5–6, 12, 15, 22, 24, 26, 32, 38, 41, 89*fn*, 148, 204*fn* 9, 220, 233, 241, 289, 289*fn*, 290, 309*fn*
 “A Puzzle about Belief,” 220–224

 language, 246, 248, 251, 286, 289–291, 301, 305–306
 metalanguage, 45–48, 51, 55–57, 59
 object language, 45–46, 56
 of thought, 248
 Lauben, Dr. Gustav, 206
 L-determinate designator, 13
 Leverage Account, 165–166, 169–172, 178, 180–182
 Lewis, David, 5, 17, 87, 88, 89
 liar, paradox of, 18, 93. *See also*
 intensional Liar paradox
 locutionary content, 192–195

- logic
 - classical, 61
 - of demonstratives, 21
 - epistemic, 13, 18
 - higher-order, 20, 86
 - intensional, 12, 13, 21
 - and metaphysics, 86–87, 90
 - modal, 12, 13, 18, 21, 22, 85–92
 - temporal, 21
- logical constant, 76
- logical form, 273, 300, 302
- logical positivism, 17
- logic of sense and denotation, 14, 15, 30, 90*fn*. *See also* Frege-Church approach to intensional logic
- Marcus, Ruth Barcan, 17
- marking (type-identification), 248–252, 256, 259–260, 264–265, 269–270, 273, 276–279
- Mate's puzzle, 132
- McGee, Van, 139
- meaning and reference, 51–52
- Michotte, Albert, 297*fn*
- modal metaphysics, 6
- mode of presentation, 250, 253, 306
- Montague, Richard, 3, 5, 7, 8, 13, 16, 20, 21, 24, 26, 28–30, 46, 159
 - on the Surprise Examination Puzzle, 18–19
- Moore's paradox, 132
- Myhill, John, 87*n*.
- naive possible worlds theory, 93, 96–98
- naive realism, 246
- name(s)
 - logically proper 34
 - vivid 35, 36*fn*
 - See also* proper name(s)
- negation, 139, 146
- neo-Kantian tradition, 247
- nonextensional context, 45, 50
- Normore, Calvin 246*fn*, 274*fn*
- norms
 - epistemic, 249
 - representational, 249, 279–280, 295–297, 308
- notional belief. *See* belief, notional
- numbers, representation of, 312–316
- oblique, 38
 - context(s), 12, 32
- occurrence
 - primary, 34
 - secondary, 34
- "Opacity," 33, 36, 38
- opaque contexts, 173
- open sentence, 33–34
- operator, 263, 312
- Oppenheim, Paul, 19
- ordinary language, 4–5
- paradox. *See* knower, paradox of; liar, paradox of; policeman paradox
- "A Paradox Regained," 159
- Parsons, Terence, 16, 160, 173, 183
- particular, 248, 252–256, 259–262, 265, 269–272, 276–81, 283–297, 305–307
- Peacocke, Christopher, 139
- Pejoratives, 139–155
- perception (perceptual representation), 246–251, 251–258, 261, 263, 265, 272–288, 251–297 *passim*
- perception-based representation, 252, 254–256, 281, 251–297 *passim*
- perceptual constancy, 250
- perspective, 246–247, 249–252, 256–257, 271, 284–286,
- Peters, Stanley, 151
- phenomenology (qualitative elements), 249–250, 295–296
- plan, 190, 192–194, 197–199
- plural representation, 253–256, 262, 265–268, 271
- possible world, 4–5. *See also* naive possible worlds theory
- possible world semantics, 4–5, 12–3, 17–8, 30, 85–92
 - paradox about, 17–8, 85–92
- Potts, Christopher 153
- predicate (conceptual attributive), 253, 255, 262–265, 274–275, 278
- presupposition, 151
- Priest, Graham, 87
- principle of plenitude, 93, 96–97
- Prior, Arthur N., 90, 138
- pronoun, 35–36

- proposition, 12, 15–16, 36, 118–119, 236–237, 240–243, 245
 - Fregean, 238–239
 - general, 36, 38
 - singular, 33, 36–39, 241, 243–245
 - See also* ramified theory of propositions
- proper name(s), 203–204, 261–262, 275, 291. *See also* individual names
- property, 250–251, 254, 257, 261, 277–279, 281–282, 293–294
- propositional attitude(s), 32, 34, 161, 163–164, 169–171, 176–177, 184, 234, 236
 - ascriptions, 148–149, 155
 - content, 238, 240, 243
 - function, 237–238
 - operator, 237
- pseudo *de re* belief(s). *See de re* belief, pseudo
- psychological ability, 251–252, 258–272, 275–297, 307–308, 311–316
- psychological explanation, 248–52, 257, 275–281, 283–284, 288, 292, 294–295, 297–298, 307–308
- Putnam, Hilary, 5, 32, 153, 238, 241, 304
- puzzles of intensionality, 98
- Pylyshyn, Zenon, 282–283*fn*
- quantification (quantifier, quantified sentence), 32, 114–120, 128–130
- quantifiers
 - adverbial, 60
 - binary, 62
 - classical, 62
 - conservative, 64
 - domain of, 62
 - expressive power, 79
 - first-order, 79, 80
 - generalized, 60
 - global properties of, 78
 - interrogative, 65
 - local properties, 62
 - sortally reducible, 63
- “Quantifying In,” 7, 33, 35–38, 38*fn*, 39*fn*, 40, 159, 171, 210–211, 233, 248, 305, 313*fn*
- quantify(ing) (quantification) into, 32, 34, 36, 38, 234–235, 240
- Quine, Willard v. O., 4, 9, 16, 22, 24, 29, 34, 36–38, 36*fn*, 39*fn*, 116, 171–172, 233–236, 238–242, 244–245, 267*fn*, 287*fn*, 305, 306
- ramified theory of propositions, 93
- recursive function theory, 14
- refer (refers, referred, referring, reference(s)), 4, 5, 6, 7, 137–138, 143–149, 153, 160, 162–167, 169–170, 172–317, 177, 186, 189–201, 233–236, 241, 243–245, 246–252, 255–257, 262, 267, 269, 289–91, 296–297, 299–302, 307–316, 246–316
 - passim*. *See also* coreference; sense/reference
- reference
 - failure, 261, 286, 294, 299 (*see also* referential illusion)
 - purported, 252–254, 270–272, 278, 284, 288
- referent, 238–239, 245
- referential illusion, 254, 292, 307–308
- referentialism, 137–138, 140
- Reichenbach, Hans, 21, 28–29
- relational belief. *See* belief, relational
- relations, representation of, 257, 261, 287, 293, 297, 306
- relevant logic, 19
- represents_B (representing_B, representation_B), 35–37, 38*fn*, 39
- representation, 247–252, 246–316
 - passim*. *See also* referential illusion
 - context-bound, 260–261, 284–285 (*see also* ability particularity)
 - singular, 246–247, 246–316 *passim*
- representational content, 247–252, 276, 283–286, 312, 246–316 *passim*
- representational function, 249, 252–255, 258, 262–263,

- 268–269, 276–281, 284–286,
288–289, 296–297, 303
representatum, 251, 277
Rescher, Nicholas, 22
Richard, Mark, 211. *See also* Russellian
Annotated Matrices (RAMs)
road back (from reference to meaning),
56–59
Rumfitt, Ian, 139
Russell, Bertrand (Russellian,
neoRussellian), 6–7, 29, 31, 34,
36–39, 41, 63–64, 87*fn*, 88–90,
110, 131, 237–238, 243, 246–247,
251–252*fn*, 261*fn*, 286, 305–306,
305*fn*, 311, 314. *See also* Russellian
intensional logic; Type Theory
Russellian Annotated Matrices
(RAMs), 211
Russellian intensional logic, 21, 89–90
Russell-Myhill paradox, 95–96
- S5 modal logic, 22, 85, 86*fn*
Salmon, Nathan, 21
Salmon, Wesley, 19, 24
scope, 34
Scott, Dana, 5, 20
Searle, John R., 297*fn*
Segerber, Krister, 5
self-attribution, 310–311, 314
self-knowledge. *See* self-attribution
semantic
 connection 126–130
 difference (SD), 110–113, 122–123,
 125
 relationism 109, 122–131, 132
 restricted plurality, 262, 265–268
 (*see also* plural representation)
 role, 109–117, 122–123, 125–126,
 130
 sameness (SS), 110–112, 122–123
 singularity (*see* generality, semantic),
 262–272, 276–278, 306–307,
 246–316 *passim*
 value, 112–115, 118–119, 125–128,
 132
semantically innocent, 47, 57
semantical system, 13
semantics
 extensional, 118–119
 intensional, 119
 sense, 37, 53, 57–58, 148, 160–163,
 168, 173, 178, 183–185
 hierarchy of 52–57
 See also Fregean sense
 sense/reference, 118–119, 131
 sense, indirect, 160, 173, 183
 singular proposition. *See* proposition,
 singular
 singular reference. *See* representation,
 singular
 singular representation. *See*
 representation, singular
 Skyrms, Brian, 18, 24
 Sorensen, Roy, 154
 sortal, 252, 258, 272, 275, 287, 291
 Stalnaker, Robert, 5, 151
 standard name, 13, 15–16, 21
 Stanley, Jason, 143
 stereotypes, 153
 Strawson, Peter F., 4, 151, 287*fn*, 307*fn*
 subitizing, 313, 316
 supervaluationism, 143
 Surprise Examination puzzle, 18–19
 syntactic singularity, 262–272, 278,
 282–287, 288–304. *See also*
 generality, syntactic
- Tarski, Alfred, 15, 19–20, 49–50, 109,
 112, 153–154
Tarski's Schema, T, 14–6, 19
Tarskian approach (scheme, etc.), 109,
 112–113, 130
term
 closed, 115
 open, 114–115
 “that,” 187–188, 196–201
 that-clause(s), 45, 49–51, 54, 56, 58
 think (thinking) about or of, 193–196,
 198–199
 thinking, way(s) of, 161–162, 165,
 169–171
 thought (Fregean thoughts), 160–178,
 180–182, 184–186
 thought (propositional), 247–252,
 253–255, 260–261, 263,
 272–275, 290–291, 296–304,
 246–316 *passim*
 tone, 148–149
 tonk (connective), 138–139
 translation, 174–178, 186

- truth, 15
- truth (propositional veridicality), 249, 262, 285
- truth condition, 285–286
- truth of, 258, 261–262, 269, 285–286, 290–291
- Trotsky, Leon, 87
- type theory, 21–22
- typographic difference, 113, 130
- understanding, 141–143, 148, 152–155, 300–304, 309, 311, 313–316. *See also* comprehension
- unique characterization, 139, 144
- unsaturated contents, 263
- vagueness, 154
- validity, 153
- van Fraassen, Bas, 5
- variable, 34, 36
 - bound, 117, 125, 129
 - free, 37, 113–114, 117, 120, 122, 125, 129
 - implicitly bound, 115–116
 - semantic role, 109–117, 122–123, 125–126, 130
 - semantic rule for, 130
 - See also* antinomy (puzzle) of the variable
- variable binding, 114–117, 119, 121, 129
 - lambda operator, 52
- variable-free, 122
- veridicality, 253, 257, 262–263, 267–269, 276–277, 283–284, 286, 288–293, 293–297, 299–304, 308
 - condition, 253, 262, 276
 - perceptual (accuracy), 253, 262, 276, 283–284, 293–297, 302, 308
 - propositional (*see* truth)
- Vienna Circle, 29
- vividness, 305–306, 314
- Vlach, Frank, 5, 21
- warrant, applicational, 289, 302
- warrant, apriori. *See* apriori warrant
- warrant, empirical, 301–303, 312
- warrant, epistemic, 248–249, 251, 289, 297–304
- way(s) of thinking. *See* thinking, way(s) of
- what is said, 189, 191–192, 194–195