REPRESENTING THOUGHTS AND LANGUAGE

David Sosa

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PHILOSOPHY
Abstract

These three papers, each constituting a chapter, lie at the intersection of philosophy of mind and philosophy of language. Chapter 1 reviews and reassesses Kripke’s puzzle about belief. I argue, contra Kripke, that the puzzle shows Millianism to be inadequate (if not incorrect). It must be supplemented with a Fregean theory. But Millianism and Fregeanism need not be opposed. Developing a distinction between mental representation and linguistic representation, I divide the notion of proposition. It is one thing to be the object of a propositional attitude, another to be the meaning of a declarative sentence. This perspective on entrenched difficulties in philosophy of mind and language discloses a unified theory accounting for all relevant intuitions.

Chapter 2 discusses externalism in the philosophy of mind. I investigate the relation between the issue of Fregeanism and Millianism in philosophy of language and the issue of internalism and externalism in philosophy of mind. Having characterized externalism precisely, I argue that it faces a serious difficulty: it cannot correctly maintain the distinction between ignorance and incoherence. Mere empirical ignorance will be enough to make an agent logically incoherent. Worse, if externalism is correct, internal duplicates might differ with respect to logical coherence.

In Chapter 3 I defend the Russellian conception of names from Kripke’s modal argument. I offer a way to accommodate Kripke’s claims about the rigidity of names without abandoning the Russellian theory.
# Table of Contents

Abstract ................................................................................................................................... iii

Table of Contents ......................................................................................................................... iv

Acknowledgements ....................................................................................................................... vi

Introduction ................................................................................................................................... 1

Chapter 1: Where is the Puzzle About Belief? ................................................................................. 9

   Part I .......................................................................................................................................... 10
      The traditional argument and its analogs .............................................................................. 10
      A tempting response rejected ............................................................................................ 19
      The hermeneutic principle .................................................................................................... 25
      Turning the tables ............................................................................................................... 29
      Development and defense of a position .............................................................................. 32
      A challenge and reply .......................................................................................................... 40
      Intermediate conclusions .................................................................................................... 47

   Part II ....................................................................................................................................... 51
      A hypothesis .......................................................................................................................... 51
      Strengthening the hypothesis ............................................................................................... 53
      Reinventing a relation ........................................................................................................... 58
      Applications to traditional puzzles ..................................................................................... 60
      Associated conceptions of logic ........................................................................................... 67

Chapter 2: Irrationality in Externalism ......................................................................................... 73

   Part I: Two Controversies ......................................................................................................... 74
      What is externalism? ............................................................................................................ 74
Acknowledgements

Mark Johnston and Scott Soames were advisers than which none better can I conceive. Since my first year at Princeton, Mark guided my philosophical development. His demand for significant philosophy challenged me; his support sustained me; his excellence was a model. Scott’s generosity during my dissertation effort was tremendous: numerous, lengthy, philosophical discussions with him transformed my thinking. Because he devotes as much energy to understanding as to responding, he is a very special critic.

While I worked on this dissertation, my life was in different ways but always powerfully and positively affected by Paul Benacerraf, Gabriela Diaz, Gilbert Harman, Jaegwon Kim, Christian Piller, Jim Pryor, Marcelo Sabates, Walter Sinnott-Armstrong, Sarah Stroud, and Elisabeth Zöllner.

My debt to Saul Kripke will become more evident below. His work is so rich that it repays (with interest) any amount of consideration.

Work on this dissertation was supported by a fellowship from the Mrs. Giles Whiting Foundation.

Special thanks go to my parents and brother Adrian—they are the bedrock upon which my life is founded—and to Gina Shishima—without her not.
Introduction

This dissertation consists of three papers, each constituting a chapter. The subject matter of the three chapters lies at the intersection of philosophy of mind and philosophy of language. I am interested in the notions of content and proposition. These notions will be central in any theory of the attitudes and they are related to other notions such as meaning, semantic value, reference, and information.

What is the significance of this range of notions? The basic philosophical item to which they are anchored is representation. Representation is important because it enables us to interact with the world in a way that would be impossible without it. If all we had were our thoughts, we would be missing something valuable. We care about connecting to the world. And we connect to the world not only by being physically a part of it, but also by representing it. Thoughts and language are impressive in their ability to be about other things. Animals with impoverished systems of representation, especially pre-verbal animals, have a less rich interaction with the world around them; they are limited to brute physical or causal relations.

Thoughts and language, because they represent, enable “presence in absence.” Loved ones living far away can be “with us” in our thoughts. We can remember relatives now dead, anticipate an important upcoming event, long for someone we miss. We can use language to recount fantastic stories, true, false, or fictional. This ability to be about other things, to make things present in absence, the exclusive province of representation, is the phenomenon of intentionality. Intentionality is a deep philosophical issue. What is the nature of representation? How does one thing get to be about another?
I do not think intentionality is a *brute* fact. It seems to me that we can exploit other relations in order to understand the relation between a thought or a bit of language and its object. Most generally, these other relations fall into two categories: instantiation and causation. It is useful, I think, to see theories of intentionality as explaining the phenomenon, ultimately, in terms of one or another (or both) of those two relations.

One aim of my dissertation is to develop and apply a general distinction between *mental representation* and *linguistic representation*. Is there a difference between what *we have in mind* when we are hopeful, afraid, expectant, or have a belief or desire and what *we say* when we produce an utterance? Is the basic business of a theory of meaning to represent the information language can communicate or is it rather to account for what we express when we “speak our mind”? According to me, a traditional understanding of *proposition* conflates two kinds of representation. Distinguishing between them and understanding how they underlie diverging trends in the theory of meaning allows a new perspective on entrenched difficulties in philosophy of mind and language.

Mental representation is independent of linguistic representation. Paradigms of mental representation can be found in phenomena such as musing, meditation, and reverie. A linguistic expression of a propositional attitude will be associated with the attitude’s mental representation. But this does not exclude the possibility that the expression exemplifies also another kind of representation—a kind of representation that the expression derives from factors (*i.e.* its linguistic context) that are independent of the belief it expresses. This, I claim, is a surprising fact: *the expression of a belief can exhibit a kind of representation that is independent of the belief it is expressing*. The nature of the relationship between mental and linguistic representation is more complex than might at first be thought.
One issue to which I apply the distinction (indeed, in the context of which the
distinction arises most clearly) is a central dispute of contemporary philosophy of
language. What does a proper name contribute to propositions expressed by sentences in
which it occurs? The leading positions derive from views advanced by Mill (and Russell)
on one side and by Frege (and Russell) on the other. Millians and contemporary neo-
Russellians say that a name is non-connotative. It contributes only its referent and has no
other semantic value. Frege (and, at times, Russell) rejected this view, arguing that a
name has another more fundamental semantic value, a sense, which is a mode of
presentation of the name’s referent.

Beginning most notably with S. Kripke’s Naming and Necessity, Fregean views
have come under sustained critical pressure. Many today take Kripke to have effectively
refuted the Fregean view of names. Later, in “A Puzzle About Belief,” Kripke defends
Millianism from a traditional objection. In response to that paper, I develop a new view
of the issue between Millianism and Fregeanism.

First, I review the (perhaps critically neglected) methodology of Kripke’s “A
Puzzle About Belief.” An investigation of exactly where the puzzle is supposed to lie
yields insight into its resolution. I argue that although Kripke is right that Millian
substitutivity is not essential to produce the puzzling results (which suggests, given other
assumptions, that Millianism is innocent), absent substitutivity some other Millian
principle is needed in order to derive the puzzle. If Kripke’s disquotation and translation
principles are inessential, as I contend, then, since the puzzle cannot be created without
some Millian principle or other, Millianism stands indicted after all.

If Millianism cannot escape arguments traditionally adduced against it, and if
Fregean theories are subject to the compelling critique mounted against them, the
discussion seems stalled and at an impasse. My proposal aims to help us escape this
apparent impasse. The distinction between mental and linguistic representation enables an interesting resolution: Fregeanism and Millianism are not, despite appearances, incompatible theories of the same phenomenon but are actually independent theories of distinct phenomena. Their apparent opposition is a matter of wholesale intertheoretic equivocation. This achieves a compromise: a unified theory of meaning that accounts for all relevant intuitions can incorporate both theories. The assumption that there is just one kind of thing that most deserves to be called “proposition” may just be mistaken; Fregeanism and Millianism should be regarded as theories of the two kinds of things that are equally good candidates.

Some of the plausibility of my proposal thus derives from its consequences for puzzles in the theory of meaning: by combining Fregeanism and Millianism in the way I propose, we gain the virtues of each and deflect the criticisms. The puzzle about belief shows that the Millian view of names is inadequate (if not incorrect). It is a theory narrowly of linguistic representation. We are left wanting a theory of meaning that sustains plausible claims about the status of an agent’s thoughts: is the agent subject to logical censure?

The notion of mental representation (as distinct from linguistic representation) is closely connected to the notion of rationality. If an agent’s rationality is to be seen as a function of relations among the contents of her thoughts, then we need something like a Fregean theory. Moreover, if Fregeanism is seen specifically as a theory of mental representation, it can redirect many standard criticisms—these criticisms derive from intuitions concerning linguistic representation. I argue that Fregeanism, understood as a theory of mental representation, survives its alleged demise.

In Chapter 2, I continue the discussion of the relation between rationality and theories of content. Similar issues arise in the new context of the controversy between
internalism and externalism in contemporary philosophy of mind. First, I discuss the relevant views themselves: what is externalism? How is it best characterized? What is the relation between the issue of Fregeanism and Millianism in philosophy of language and the issue of internalism and externalism in philosophy of mind? This relation is shown to be more complex than might appear. I show, for example, that Millianism is compatible with internalism. The key point is that while Fregeanism and Millianism differ over the constitution of propositions, internalists and externalists differ over the determination of the propositional content of attitude states.

After clarifying the relation between the two controversies, I redeploy a question that lies behind some of the discussion in Chapter 1: what constraints are imposed on a theory of propositional content by the notion of rationality? By the end of Chapter 1, I will have argued that Millianism lacks the resources to explain an agent’s rationality in terms of contents of thought. In Chapter 2, I argue that externalism faces an analogous problem. By making the contents of an agent’s thoughts a matter that is not determined by her overall internal state, externalism will blur the distinction between ignorance and incoherence. A little bit of empirical ignorance will suffice to make an agent logically incoherent. This empirical ignorance will be the product of circumstance, literally, a matter of relations between an agent and his environment. Agents who are logically innocent and fully rational, but who have false empirical beliefs, may be mischaracterized as prone to logical error.

Moreover, if externalism is correct, then it will be possible for two agents who are internal duplicates to differ with respect to logical coherence and thus with respect to rationality. One of two internal duplicates may be inclined to draw what according to externalism is an invalid inference while the other is not so inclined. It seems to me, however, that one thing we know about logical coherence is that it’s an internal matter. Duplicates are coherent, or not, as a pair.
In light of my earlier reaction to the analogous situation in philosophy of language, the possibility arises that internalism and externalism are not so much incompatible theories of the same phenomenon as independent theories of different phenomena. An idea to be investigated further, for example, is whether the kind of content that ‘ain’t in the head’ is just the proposition that would be expressed by an ordinary assertoric utterance. In other words, it might be that internal duplicates would share all and only the same mental representations and differ, potentially, only with respect to linguistic representations. I have not attempted to extend my position in that way here; but it is an attractive and natural extension.

Chapter 3 defends Russell’s claim that names are disguised definite descriptions. I consider critically Kripke’s attack on that claim; in particular, I confront the so-called modal argument. Why focus so narrowly on one argument by one philosopher against the view of another? Kripke’s argument—though, by his own admission, not stated thoroughly or in detail—has been widely influential. Many today take Kripke (together with Donnellan and Putnam and others) to have decisively refuted the Russellian view of names. The argument on which I will focus may seem to be the most decisive. It is interesting too because of its connection with Kripke’s thesis of names as so-called “rigid designators”—a thesis for which he is responsible and celebrated. Also, I believe that Russell’s view contains deep insights into the nature of the referential or intentional relation, insights which are less well appreciated today. Only by dispelling the widespread feeling that Russell’s theory is inadequate to actual linguistic data can we begin to understand again some of Russell’s deeper philosophical motivation.

The root of my proposal is that the truth conditions, with respect to a counterfactual situation S, for a simple sentence are the truth conditions of a complex sentence that embeds the simple sentence under the modal operator “with respect to S.” Since according to Russell the complex sentence may be ambiguous, there may be
alternative truth conditions, with respect to a counterfactual situation S, even for a simple sentence. If all this is right, the intuitions about the rigidity of names do not so much rebut as supplement the Russellian theory of names. My proposal does not reject Kripke’s claims about the rigidity of names; I find a way to accommodate those claims within a broadly Russellian theory. Residual differences between names and definite descriptions can be explained in terms of conventional pragmatic restrictions.

My dissertation was written during what can seem like a crisis in the philosophy of mind and language. In particular, with respect to issues concerning content and proposition (what is their nature, what is the proper analysis of attitude ascriptions, what determines the contents of an agent’s attitudes) there appears to be argumentative stagnation. Millians and neo-Russellians disagree with Fregeans; but neither side seems able to make a significant impact on the other. The situation is similar with respect to the dispute between externalism and internalism (though externalism is clearly the predominant position). I hope the ideas I defend can help us to understand why that is so, while properly crediting all sides with true insight.
Chapter 1: Where is the Puzzle About Belief?

Rationality is a matter of having one’s beliefs in order. So if there is a puzzle about belief, there is a puzzle about rationality. In fact, I think Kripke’s famous puzzle about belief relies essentially on intuitions about an agent’s rationality. Here is one passage from the article:

...We may suppose that Pierre...is a leading philosopher and logician. He would never let contradictory beliefs pass. And surely anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them. Precisely for this reason, we regard individuals who contradict themselves as subject to greater censure than those who merely have false beliefs. But it is clear that Pierre...is in no position to see, by logic alone, that at least one of his beliefs must be false. He lacks information, not logical acumen. He cannot be convicted of inconsistency: to do so is incorrect. (Kripke, 1988, p. 122)

In spite of the connection Kripke establishes, relations between rationality and theories of propositional attitudes have been underappreciated, it seems to me. I here return to Kripke’s puzzle with special attention to the role of rationality.

In the first part of this paper, my investigation will focus on three areas. First, in what does the puzzle consist, exactly? A second focus will be the methodology of Kripke’s paper. I believe the puzzle and Kripke’s use of it as a defense of Millianism involves an interesting methodology which has not been given the attention it deserves. My own rejection of some of Kripke’s position will involve the same maneuver he uses to defend Millianism. Third, I will discuss relations between rationality, meaning, and reference, as they arise in connection with the puzzle. Appreciating these connections renders Fregean theories more attractive. In the second part of the paper, I consider the possibility that there is more than one notion of meaning involved. In connection with
this general discussion, I will introduce a general distinction between ways of representing and will argue that understanding this distinction helps make sense of philosophical issues about intentionality in philosophy of mind and language.

Part I

The traditional argument and its analogs

In “A Puzzle About Belief,” Kripke is concerned to cast doubt on the conclusion that belief contexts are not ‘Shakespearean.’ Geach introduced the name ‘Shakespearean’ after the line, “A rose / By any other name / would smell as sweet.” What makes a context Shakespearean? Being Shakespearean is just being transparent with respect to proper names: the context may still be opaque with respect to other referring expressions (such as definite descriptions). Kripke does not claim to establish that belief contexts are Shakespearean: ultimately he thinks that it is simply puzzling what the truth about certain examples might be. The moral of his paper, according to Kripke, is that we are not in a position to draw any definite conclusions about substitutivity in belief contexts.

Shakespeareanism is constituted by the following principle, given first informally and then with a substitutional schema (the substitution class for ‘N’ and ‘N’’ is the class of proper names. Also, with Kripke (p. 104) I confine myself throughout to expressions that do not involve quotation or logophora

(S)hakespeareanism Codesignative proper names are intersubstitutable in other expressions salva veritate,

or,

∪...N...' & N=N' □ ∪...N'...'.

1 E.g. expressions such as “Giorgione was so-called because of his size.”
Shakespeareanism follows from a philosophical position concerning the semantic value of proper names.\(^2\)

\[(M)\]illianism The only semantic value of a name is its reference.

Kripke believes that a certain traditional argument against Millianism fails. This traditional argument exploits the fact that \((M)\) implies \((S)\), together with an apparent reductio of \((S)\), in order to reject the position defined by \((M)\). Kripke claims that the apparent reductio of \((S)\) is a mere appearance; the same contradictory result can be derived, in a parallel fashion, without the use of \((S)\).

So Kripke does not attack the traditional argument directly; rather, he tries to recreate, without invoking \((S)\), the contradictory results. To this end he introduces two very basic principles. The first is:

\[(D)\]isquotation\(_{\text{English}}\) If a normal English speaker, on reflection, sincerely assents to ‘p,’ then he believes that \(p\).

With respect to this principle Kripke offers the following clarificatory note: “The sentence replacing ‘p’ is to lack indexical or pronominal devices or ambiguities that would ruin the intuitive sense of the principle” (pp. 112–3, emphasis mine). ‘\(P\)’ is assumed to be a sentence of English. We can imagine analogous disquotation principles for other languages—\((D)\)isquotation\(_{\text{Spanish}}\), \((D)\)isquotation\(_{\text{French}}\), etc.—written in that language, in which ‘English’ is replaced by a name of that language (also in that language) and where ‘p’ is assumed to be a sentence of that language. The second principle is:

\(^2\) Actually, an even stronger substitutivity principle which replaces “\(\text{salva veritate}\)” by “\(\text{salva significatione}\)” also follows from the Millian position to be defined.
If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language).

Importantly, Kripke sees the traditional argument against Millianism as itself implicitly involving (D). He might render the traditional argument thus (keep in mind that the renditions I give here and below are meant to be not formal proofs, but schematic representations of how the examples are supposed to yield their results):

**Kripke’s rendition of the traditional argument**

1. Jones is rational. Assumption
2. Cicero = Tully Assumption
3. Jones assents to “Cicero denounced Catiline.” Assumption
4. Jones assents to “Tully did not denounce Catiline.” Assumption
5. Jones believes that Cicero denounced Catiline. 3, (D)English
6. Jones believes that Tully did not denounce Catiline. 4, (D)English
7. Jones believes that Cicero did not denounce Catiline. 2, 6, (S)
8. Jones believes that Cicero denounced Catiline and Jones believes that Cicero did not denounce Catiline. 5, 7, conj.
9. If Jones believes that Cicero denounced Catiline and Jones believes that Cicero did not denounce Catiline, then Jones has contradictory beliefs. ?
10. Jones has contradictory beliefs. 8, 9, m.p.
11. If Jones has contradictory beliefs, then Jones is not rational. Def. ‘rational’

---

3 If this seems too strong, remember that the beliefs in question are all held occurrently and reflectively (we are not concerned with dispositional or implicit belief) and that the agent in question can be assumed to be “a leading philosopher
Discharging our assumptions, we may seem to derive the negation of (S) since it seems to be the most suspect principle used. (M) implies (S), so we seem to derive the negation of (M) by contraposition. At least that, as Kripke seems to see it, is how the traditional argument is supposed to work.

Kripke will show that this argument, by itself, is insufficient. He will recreate the same problematic result using the same assumptions (together with other apparently indubitable principles) used in his rendition of the traditional argument. In the traditional argument, that problematic result is turned against (S). But in Kripke’s analogous arguments, (S) is not used. The only substantive principles used are, it seems, the very basic principles of disquotation and translation.

The first of Kripke’s analogs to the traditional argument begins with a story about Pierre, a normal French speaker who lives in France but has heard of that famous distant city, London (which Pierre of course calls ‘Londres’). On the basis of what he has heard of London, he is inclined to say, in French, “Londres est jolie.” We can conclude that he believes that London is pretty. Later, Pierre moves to an unattractive part of London. He learns English by the ‘direct method,’ without using any translation between English and French. Pierre is unimpressed with his surroundings and is inclined to assent to the English sentence “London is not pretty.” We can conclude that he believes that London is not pretty. But then Pierre believes that London is pretty and Pierre believes that London is not pretty. This means that Pierre has contradictory beliefs. However, it is clear that Pierre lacks information, not logical acumen. He should not be convicted of

inconsistency. This is Kripke’s presentation of what I’ll call the ‘(unsimplified) bilingual analog.’ (See Kripke, p. 122.)

The (unsimplified) bilingual analog

1. Pierre is rational. Assumption
2. Pierre, en réfléchissant, donne son assentiment à “Londres est jolie.” Assumption
3. Pierre, on reflection, assents to “London is not pretty.” Assumption
4. Pierre croit que Londres est jolie. 2, (D)French
5. Pierre believes that London is pretty. 4, (T)
6. Pierre believes that London is not pretty. 3, (D)English
7. Pierre believes that London is pretty and Pierre believes that London is not pretty. 5, 6, conj.
8. If Pierre believes that London is pretty and Pierre believes that London is not pretty, then Pierre has contradictory beliefs. ??
9. Pierre has contradictory beliefs. 7, 8, m.p.
10. If Pierre has contradictory beliefs, then Pierre is not rational. Def. ‘rational’
11. Pierre is not rational. 9, 10, m.p.

The analogy with the traditional argument is evident. In both cases we have a presumably rational agent who, according to certain seemingly innocent principles, is characterized as inconsistent and subject to serious epistemic censure. The traditional argument against Millianism concludes that the seemingly innocent principles are not—that (S) is guilty. But Kripke’s analog proceeds independently of (S).
Another, simpler representation of Kripke’s bilingual analog to the traditional argument could be created using a generalization of (D).

\[(D')\text{isquotation}\] If a normal L-speaker, on reflection, sincerely assents to ‘p’ (a sentence of L), and if ‘q’ is a translation (into English) of ‘p,’ then S believes that q.\(^4\)

The (simplified) bilingual analog

(1) Pierre is rational. Assumption
(2) Pierre, on reflection, assents to “Londres est jolie.” Assumption
(3) “London is pretty” is a translation (into English) of “Londres est jolie.” Assumption
(4) Pierre, on reflection, assents to “London is not pretty.” Assumption
(5) Pierre believes that London is pretty. 2, 3, (D’)
(6) Pierre believes that London is not pretty. 4, (D’)
(7) Pierre believes that London is pretty and Pierre believes that London is not pretty. 5, 6, conj.
(8) If Pierre believes that London is pretty and Pierre believes that London is not pretty, then Pierre has contradictory beliefs. ?
(9) Pierre has contradictory beliefs. 7, 8, m.p.

\(^4\) Berel Lang has suggested (in conversation) that these disquotation principles seem to be subject to the following kind of counterexample: A normal Spanish speaker might, on reflection, sincerely assent to ‘Esto es una oración en Español.’ ‘This is a sentence of Spanish’ might be the translation (into English) of ‘Esto es una oración en Español.’ But it still might be false that the Spanish speaker believes that this is a sentence of Spanish (for starters, it isn’t a sentence of Spanish). It seems to me, however, that Kripke meant to exclude such examples with his limitation of disquotation principles to sentences which “lack indexical or pronominal devices…that would ruin the intuitive sense of the principle” (1988, pp. 112–3).
(10) If Pierre has contradictory beliefs, then Pierre
    is not rational.               Def. ‘rational’

(11) Pierre is not rational.          9, 10, m.p.

An advantage of this simplified representation may be that every line of the
argument is in the same (English) language.\(^5\) I do not think, however, that there is any
substantial problem with the fact that in the unsimplified bilingual analog, the left side of
lines (2) and (4) are in French. The argument is given only informally anyway and that
feature could presumably be eliminated if the argument were represented formally.\(^6\) In
relevant respects, the unsimplified and simplified bilingual analogs are equivalent.

Now, in his first analog, Kripke uses both the disquotation and translation
principles. Only (D), however, is plausibly taken to be implicit in the traditional
argument. I believe (T) is partly definitory of translation and is thus analytic. Notice
what a weak principle (T) really is: it establishes only a weak necessary condition on
translation. It claims only that translation must preserve truth value. This can, but should
not be, confused with the obviously false claim that translation need only preserve truth
value. In case there is any temptation to suspect (T), Kripke creates another analog which
makes no appeal even to (T) and does not employ (D').

This second example involves Peter, who learns the name ‘Paderewski’ with an
identification of the person named as a famous pianist. Naturally, Peter assents to
“Paderewski had musical talent” and we can infer that Peter believes that Paderewski had
musical talent. Later, in a different circle, Peter learns of someone called ‘Paderewski’

\(^5\) We also avoid the use of (T) and with it Kripke’s claimed need for a form of the
Tarskian disquotation principle for truth (see Kripke, 1988, p. 142n26).

\(^6\) Exactly how such a formal representation should relate to an informal
representation is, nevertheless and as will be emphasized below, a very significant
issue.
who was a Polish nationalist leader. Peter is sceptical of the musical abilities of politicians. Thus, Peter assents to “Paderewski had no musical talent.” By parallel reasoning, we would infer that Peter believes that Paderewski had no musical talent.

The monolingual analog

(1) Peter is rational. Assumption
(2) Peter, on reflection, assents to “Paderewski has musical talent.” Assumption
(3) Peter, on reflection, assents to “Paderewski does not have musical talent.” Assumption
(4) Peter believes that Paderewski has musical talent. 2, (D)English
(5) Peter believes that Paderewski does not have musical talent. 3, (D)English
(6) Peter believes that Paderewski has musical talent and Peter believes that Paderewski does not have musical talent. 4, 5, conj.
(7) If Peter believes that Paderewski has musical talent and Peter believes that Paderewski does not have musical talent, then Peter has contradictory beliefs. ?
(8) Peter has contradictory beliefs. 6, 7, m.p.
(9) If Peter has contradictory beliefs, then Peter is not rational. Def. ‘rational’
(10) Peter is not rational. 8, 9, m.p.

Having produced these analogs, Kripke claims they show that the traditional argument should not be seen as a reductio of (S). According to him, the use of Jones’s case as a counterexample to (S) is analogous to the following situation:
Someone wishes to give a *reductio ad absurdum* argument against a hypothesis in topology. He does succeed in refuting this hypothesis, but his derivation of an absurdity from the hypothesis makes essential use of the unrestricted comprehension schema in set theory, which he regards as self-evident. (In particular, the class of all classes not members of themselves plays a key role in his argument.) Once we know that the unrestricted comprehension schema and the Russell class lead to contradiction by themselves, it is clear that it was an error to blame the contradiction on the topological hypothesis. (p. 118)

It is Kripke’s suggestion that the situation with respect to the interchangeability of codesignative proper names is similar. How should we respond to this suggestion?

My own response is to accept the point that the traditional argument, by itself, is insufficient to indict (S). But I believe that the traditional argument can still be used to rebut (M). Thus I accept Kripke’s central point but deny his more general moral. Right now it may be puzzling how, if I accept Kripke’s point that (S) is innocent, I expect to rebut (M). Much more will be said below in defense of my position. First, however, I will support my acceptance of his central point by considering and rejecting a response that denies that central point.

A tempting response rejected

According to one tempting response to Kripke’s article, even if an analogous contradiction can be created using other principles, this does nothing to show that the traditional argument is not a reductio of (S). Since in both analogs some kind of

---

7 Of course the topological hypothesis, and (S), may nevertheless be false. Kripke’s claim is simply that if the situation is as he says, we should not yet *claim* that they are false.
The defender of the current response denies Kripke’s claim that assumptions (3) and (4) in the traditional argument are implicitly derived from the deeper assumption of Jones’s assent to corresponding sentences (as in Kripke’s rendition). According to this
response, there is a logically possible world in which a rational agent believes that Cicero
denounced Catiline and also believes that Tully did not denounce Catiline. The question
of how, in practice, we justify making such attributions is avoided—(D) is unnecessary;
we simply hypothesize a world in which the attributions are true.

This response seems to me to be partly correct but to be ultimately inadequate. It
is correct that the traditional argument need not proceed by way of any disquotation
principle. A remarkable difference between the traditional argument and Kripke’s
analogs is that the following attributions are *prima facie* plausible—S believes that
Cicero denounced Catiline and S believes that Tully did not denounce Catiline—in a way
that these other attributions are not—S believes that London is pretty and S believes that
London is not pretty. With respect to the latter attributions, though not with respect to the
former, we would immediately presume that S is not rational, or that there has been some
equivocation. So the disquotation principle seems indispensable to Kripke’s analogs in a
way that it does not seem indispensable to the traditional argument.

Thus far the current response may be right; nevertheless, I think it is ultimately
inadequate. The reasons for its inadequacy are of two kinds: (i) by claiming that Kripke’s
analogs are reductios of (D), it denies the disquotation principle (and this is a
disadvantage of any such response to the puzzle) and (ii) puzzles analogous to the
traditional argument can be generated with more obviously unobjectionable disquotation
principles or even without any disquotation principle at all. (ii) suggests that what is
essential to the puzzle about belief is independent of the disquotation and translation
principles (so that we can assume them for the sake of argument).8 Here then are four

---

8 It seems to me that Kripke himself overemphasizes the role of the disquotation and
translation principles in the creation of the puzzles.
reasons against the current response to Kripke’s puzzle (in what I take to be increasing order of strength).

First, the principle of disquotation is a relatively basic principle governing our practice of belief attribution. Whatever its merits, it is clear that we rely on it in practice. This seems to me to be some reason against the current response to Kripke’s puzzle, if not itself a decisive reason. There may be many principles upon which we rely in our practices which are philosophically questionable. If it were the case, however, that without such principles of disquotation many belief attributions that we take to be justified would not be justified, then denying those principles would leave us with the task of finding an alternative justification for those attributions. Thus it is at least a disadvantage, if not necessarily an insuperable disadvantage, of this response that it involves the denial of the principles of disquotation.

Second, as Salmon points out, “an alternative version of Kripke’s puzzle can be generated by means of a different and obviously unobjectionable disquotation principle…conditionally linking assertion to utterance of a sentence.” In other words, whether or not sincere reflective assent to a sentence in a language with which one is competent implies one’s belief in the proposition expressed by that sentence, it certainly implies one’s assertion of that proposition, and the same circle of issues arises if the puzzle is given in those terms. This suggests that we merely avoid the deep issues if we try to block the puzzle by denying the disquotation principles.

Third, to the extent that Kripke’s examples do seem to parallel the traditional examples urged against Millianism, the above response has the disadvantage that it must claim a substantial difference between the examples. In the traditional cases, it is now

---

9 Salmon, 1986, p. 130.
claimed, the contradictory results are the result of (S), while in Kripke’s cases, the culprit is some variant of (D). Notice that in the traditional argument there is not even the temptation to doubt assumption (3) or (4). Why then in Kripke’s cases prohibit the move to the analogous theses? As Kripke claims, “[i]ntuitively, Jones’s assent to both ‘Cicero was bald’ and ‘Tully was not bald’ arises from sources of just the same kind as Pierre’s assent to both ‘Londres est jolie’ and ‘London is not pretty’” (p. 133). Thus it seems that if the belief attributions in the traditional argument are acceptable, they are also acceptable in Kripke’s cases. The above response must deny this intuition.

Finally, and this seems to me the strongest point against the response under consideration, it may be that Kripke-type cases can be created even without any principle of disquotation. For this one need only imagine oneself in the position of Pierre or (better) of Peter. I used to think that it was an interesting coincidence that there was a well-known politician named John Glenn since there was also a famous astronaut by the same name (I thought the astronaut had died on a later mission). Disquotation seems to be irrelevant: Considering my situation before I discovered my error, I would attribute to myself the belief that John Glenn has been in space as well as the belief that John Glenn has not been in space. It’s not that I infer from my assent to ordinary language sentences, by means of the principle of disquotation, that I must have had those beliefs. I don’t need to infer which beliefs I had; I remember them. And among them, I claim, were ones that are properly reported as that John Glenn has been in space and as that John Glenn has never been in space.10 Perhaps you have sometime found yourself in that kind of

10 At the time, I might even have said, sincerely, that I believed that John Glenn has been in space and that John Glenn has not been in space—just as today I might say, sincerely, that I believe that banks are a good place to store money and that banks are not a good place to store money.
situation; almost certainly you can imagine yourself in that situation. The general point is that it is useless to try to stop Kripke’s examples before they reach the step in which the agent is attributed the beliefs that N is F and that N is not F.

We may even suspect that Kripke has ‘put the cart before the horse’ in deriving Pierre’s beliefs from his assent to ordinary language sentences. For if, as Kripke himself says (as we saw above), “Jones’s assent to both ‘Cicero was bald’ and ‘Tully was not bald’ arises from sources of just the same kind as Pierre’s assent to both ‘Londres est jolie and ‘London is not pretty’” [emphasis mine], then Jones’s assent and Pierre’s assent arise from some sources. What might those sources be? Their respective beliefs in what those sentences express perhaps?

The reasons I’m giving against this tempting response to Kripke’s analogs suggests that the role of (D) in the puzzle has been overemphasized. Kripke shows (implicitly) how to create analogous puzzles using the same kind of assumptions and principles used in the traditional argument, whether or not we take the traditional argument to involve disquotation. Notice that I am using Kripke’s own methodology here: since the puzzle can be recreated, ceteris paribus, without (D), we should not blame the contradictory results on that principle. Thus, the issue of (D)’s acceptability is not central in the puzzle about belief. It may be better to accept that principle if possible, but I think the central issue is independent.

---

11 In an example of Salmon’s concerning the sentence “Aristotle is Aristotle,” “though we do not know it, the philosopher of antiquity did not die in ancient times as we think, but went into hiding, discovered the philosopher’s stone which slows down the aging process, and emerged in the twentieth century as the powerful shipping magnate” (Salmon, 1986, p. 175). If this example were true wouldn’t we all be in a position like that of Pierre in our believing both that Aristotle died in antiquity and that Aristotle (the shipping magnate) did not die in antiquity (but in fact was born in and lived his whole life in the twentieth century)?
I give these reasons against the alternative possible response to Kripke sketched above. As I have said, I prefer to accept Kripke’s central claim that the traditional argument (by itself) is not a reductio of (S). But I do not believe the traditional argument is thereby rendered ineffective against Millianism. I propose to transform the traditional argument, together with Kripke’s analogs, into an attack on (M) that circumvents (S).

If I am right that neither (S) nor (D) nor (T) is to blame for the contradictory results obtained in the traditional argument and in the analogs—then what is to blame for the absurd results obtained in the puzzle cases?

The hermeneutic principle

Notice that some situations do not allow for the creation of an analogous puzzle. By seeing what would prevent the creation of a Kripke-style analog, we may understand better what is essential to those analogs. Rock is a normal monolingual English speaker who has never left his small town of Paris, Texas. Of course he has heard of that famous distant city, Paris, though he himself has never been to France. On the basis of what he has heard of Paris he is inclined to say, in English, “Paris is pretty.” On the basis of his sincere English utterance, we will conclude: (i) Rock believes that Paris is pretty. Rock’s opinion of his own town is much less favorable. Based on what he sees, he is inclined to assent to the English sentence: “Paris is not pretty.” On the basis of his sincere assent to that sentence, we should conclude: (ii) Rock believes that Paris is not pretty. The (non-)puzzle is how to describe this situation: Does Rock, or does he not, believe that Paris is pretty? We have:

The unsuccessful analog

(1) Rock is rational                                    Assumption
(2) Rock believes that Paris is pretty.                Assumption
What is the difference between this and Kripke’s original puzzles? And how should we resolve this puzzle? Of course, in this case ‘Paris’ is ambiguous. But to say this is not yet to solve the (non-)puzzle. Kripke is adamant that “[i]t is no solution to observe that some other terminology, which evades the question whether Pierre believes that London is pretty, may be sufficient to state all the relevant facts” (p. 123). We should not evade the question whether Rock believes that Paris is pretty—does Rock, or does he not, believe that Paris is pretty? Or, for that matter, do you, or do you not believe that Paris is pretty? The relevance of the ambiguity of ‘Paris’ must be made explicit.

It is especially clear that attributing to Rock both the belief that Paris is pretty and the belief that Paris is not pretty is not attributing to him contradictory beliefs. Though “Paris is pretty” and “Paris is not pretty” have syntactic structures which often translate into logical contradictories in a regimented language, in this case we use different

---

12 Steps (2) and (3) here can be derived from assumptions and a disquotation principle; but such a derivation would be redundant. The example I have created is logically possible (probably lots of people are actually in Rock’s position) and the attributions seem plausible in the example as given.
constants for the different occurrences of ‘Paris.’ Those sentences would be translated as ‘Fc’ and ‘¬Fe’.

The insight to be gained is that it is not the case for all ordinary language names Ψ that ∪Ψ is F and ∪Ψ is not F are contradictory. This contrasts sharply with formal constants: for any formal constant φ, ∪Fφ and ∪¬Fφ will be contradictory. The issue is then: when can we represent the ordinary language sentences ∪Ψ is F and ∪Ψ is not F as ∪Fφ and ∪¬Fφ, respectively. Well, it seems that Ψ must not be ambiguous.

Where does this consideration take us? In the unsuccessful analog, step (5) seems illegitimate: return now to the corresponding steps in the traditional argument and in Kripke’s analogs. In each of those arguments there was a step analogous to step (5) here. In each case I left only a question mark as the justification for that line. To do justice to our intuitions about these arguments, we must find a principle to substitute for those question marks such that the principle will seem to justify the step in the traditional argument and in Kripke’s cases but will clearly make the corresponding step illegitimate in the unsuccessful analog.

Here is the principle I believe is at work: 13

(H)ermeneutic If a name in ordinary language has a single referent then it may correctly be represented logically by a single constant.

‘Cicero,’ ‘London,’ and ‘Paderewski’ in the earlier arguments each has a single definite reference. According to (H), then, we can represent each of them logically by a

13 It may be that a stronger hermeneutic principle is really at work:

(H’)ermeneutic A name in ordinary language may correctly be represented logically by a single constant iff the name has a single referent.
single constant. Given the reasonable assumption that an agent has contradictory beliefs just in case the contents of her beliefs are themselves contradictory, we can justify, respectively, step (9) of the traditional argument and step (7) of the monolingual analog. ‘Paris’ in the unsuccessful analog does not have a single referent—this distinguishes step (5) of the unsuccessful analog. In effect, (H) takes a name’s having a single referent as sufficient for its being unambiguous (in the sense relevant to representation by a single logical constant).

Since we naturally accept those steps in the traditional argument and in Kripke’s analogs, but reject it in the unsuccessful analog, there must be a replacement for the ‘?’ in each of those steps such that the replacement seems to justify the move in the traditional argument and in Kripke’s analogs but seems not to justify the move in the unsuccessful analog. I claim that (H) is such a replacement. The problem with the unsuccessful analog is that it does not satisfy the antecedent of (H): ‘Paris’ in that example clearly does not have a single referent. In the other cases, however, the relevant constant term does satisfy the antecedent of (H). Thus (H) would justify the move in those cases.

Turning the tables

Quickly to turn the tables however—why not see those earlier arguments as so many reductios of (H)? If (H) is essential in recreating the contradictory results, that is, if an analog of the traditional argument cannot be created without the use of (H), then it would seem that Kripke’s analogs are reductios of (H). Consider the consequences of such a position.14

14 If (H) follows from (S), any reductio of (H) is a reductio of (S). Thus we may after all have an argument that the traditional argument is a reductio of (S), if only indirectly.
Importantly, recalling the general restriction to expressions not involving quotation, notice that (H) follows from (M) but need not follow from a position that allows a proper name in ordinary language to have a semantic value other than its referent. If the only semantic value of a proper name is its referent, as in Millianism, then a name’s having a single referent would be sufficient to insure the propriety of representing that name by a single logical constant.

On the other hand, by allowing a proper name to have a semantic value that determined but was not determined by its referent, one would be in a position to place stronger conditions on the representation of a name by a single logical constant. If one held that proper names had senses as well as referents, and one held that senses could vary even while the referent remained the same, then one could allow that a name is ambiguous even while it has only a single definite reference. A name that has more than one sense would then have to be represented by a number of different logical constants, one for each of the name’s senses. In short, Fregeanism can see ambiguities to which Millianism is blind.

Notice that I here abstract away from any particular user of the name. This may be an extension of Frege’s own position but seems to me to be fundamentally in keeping with his view of senses as objective, user-independent features of terms. On the view I’m considering, there may be more than one sense associated, for a single user, with a single term that has a single referent.

My suggestion is that an essential step of the analogs is justified by appeal to a principle, (H), that in turn is justified only by Millianism. The examples then amount to so many reductios of the principle and with it of Millianism (so long as (D) is not essential to those examples, as I have argued). (H) takes a name’s having a single referent as sufficient for its being unambiguous, but there may be a kind of ambiguity that is
compatible with having a single referent. If for example ‘London’ in the bilingual analog is ambiguous in this way—and it is not implausible to suppose that it is—then even if Pierre believes both that London is pretty and that London is not pretty, he may have consistent beliefs.

A natural response to what I’m suggesting here might be, [incredulous] “What? Are you claiming that even if Pierre believes that London is pretty and that London is not pretty (and we’re talking only about London, England here), he may have consistent beliefs?!” The appeal of such a response is that the claims whose justification I’m challenging seem prima facie unimpugnable. I would like to suggest that this appearance is the result only of a very tempting but ultimately controversial presupposition.

Remember earlier (p. 3) when I highlighted the word “ambiguous” in Kripke’s clarificatory comments on the principle of disquotation. In the context of a dispute between Millianism and Fregeanism the notion of ambiguity is indirectly called into question. Why? Because that dispute is a dispute over the kind of semantic value of a certain kind of term. Fregeanism postulates a kind of semantic value that determines but is not determined by the kind of semantic value postulated by Millianism, and since whether or not a term is ambiguous (in the sense relevant here) depends on whether it has more than one semantic value\(^\text{15}\) (of a given kind), the notion of ambiguity is indirectly called into question in the dispute between Millianism and Fregeanism.

\(^{15}\) Though, if it were lexicographically preferable, Fregeanism could allow ‘ambiguous’ to mean having more than one referent. Support for this would be the intuition that whether or not, e.g., Peter’s beliefs are contradictory, what he says in English is contradictory. Of course if Fregeanism allowed this terminology, it would then have to demand, for the proper representation of a singular term by a logical constant, not only that the term be unambiguous but more generally that the term not have more than a single basic semantic value. (It might also need to divorce the concept of what a sentence says from that of propositional content. See below Part II, pp. 52–74.)
Given the possibility of what might be called Fregean-ambiguity (in which a term is ambiguous in spite of having only a single definite reference), it is even unclear whether standard presuppositions of logical representation are satisfied. In particular, as we saw, even if Rock believes that Paris is pretty and that Paris is not pretty, he may not have contradictory beliefs. Of course, comes the response, since “Paris” is ambiguous in the example as given, “Paris is pretty” and “Paris is not pretty” do not contradict each other. I agree. But what is the notion of ambiguity being used?

One way to put the presupposition I am challenging is as follows: *ambiguity is a matter of multiple reference*. If it were required, for a term to be ambiguous, that it have more than one referent, then, it seems to me, Millianism would be presupposed; such a requirement excludes a Fregean position in which a name with a single referent is ambiguous in virtue of having more than one sense. And Kripke’s project is precisely to recreate a difficulty without presupposing Millianism.

Development and defense of a position

This then is substantially my analysis of Kripke’s puzzle: as a defense of Shakespeareanism from the traditional argument, Kripke’s puzzle about belief is a success as far as it goes. But appearing as it does in the larger context of the issue between Millianism and anti–Millianism, it fails adequately to defend Millianism. Millianism implies not only a very liberal principle of substitutivity, but also a very liberal hermeneutic principle. Either of these principles, (S) or (H), leads to the

---

16 It is also unclear whether Kripke’s principle of disquotation applies. By ‘applies’ I mean that its presuppositions are satisfied, not that it is true. I have argued that the issue of (D)’s truth is independent of the central issue of the puzzle about belief. Here I note, and I will develop the claim below (pp. 36–9), that the issue of (D)’s *applicability*, whether or not it is true, is not independent of that central issue.
characterization as contradictory of beliefs held occurrently and reflectively by rational agents. But no rational agent occurrently and reflectively holds contradictory beliefs.\textsuperscript{17}

For contrast, imagine a view according to which some singular terms can be ambiguous even when they have a single referent but other terms cannot. In particular, imagine a position that held that definite descriptions can be ambiguous even when they have only a single referent but that any name that has a single definite reference cannot.\textsuperscript{18} An example should help to make this position clearer. Take an expression like “the owner of the banks.” That expression might refer to a single real estate and financial tycoon. By now we can easily imagine how the puzzle could go: Johnson, who lives in a small town with only two financial institutions and one river, has an account at Acme Savings & Loan and has on occasion seen the majority shareholder of that institution who, Johnson knows, is also the majority shareholder of National Deposit & Trust across town. He is skeptical, however, of the generosity of financial tycoons and believes that the owner of these banks is not generous. Later, while fishing at the nearby river, Johnson meets the owner of the surrounding property (and doesn’t recognize her). She tells him that he is welcome to fish on her land whenever he wants. Johnson is impressed by her generosity (he believes the owner of these banks is generous). Is Johnson irrational?

\textsuperscript{17} Here I make a controversial assumption; \textit{cf.} Salmon (1986, Appendix A). But recall Kripke’s own emphasis that “anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them.” And what is undeniable is the intuition that we are \textit{more} responsible for the simple logical relations among our beliefs than for other kinds of epistemic virtue (\textit{e.g.}, proper sensitivity to evidence, overgeneralization, \textit{etc.}). I emphasize that the contradictory beliefs in question would be \textit{logically} contradictory; their form would be such that one is simply the negation of the other. No difficult deductive procedures need be employed to ascertain their consistency. And again, they are held \textit{occurrently}, not dispositionally. See also below Chapter 2, pp. 98–102.

\textsuperscript{18} It should be noted that the above example involving “Paris” in no way supports this position since in that example “Paris” does \textit{not} have a single definite reference. Its ambiguity is not in spite of a single definite reference but precisely because of the existence of more than one reference.
It is interesting to note that the Millian (as well, perhaps, as Kripke) might respond to such an example in exactly the way I have responded above to Kripke’s analogs. The example makes it plausible to assume that Johnson believes that the owner of the banks is generous and that he believes that the owner of the banks is not generous. Now, are those beliefs contradictory? In this case, Millians would say that “the owner of the banks is generous” and “the owner of the banks is not generous” do not contradict each other. “The owner of the banks” is ambiguous (in the relevant sense) even though it refers definitely to a single individual. Millians insist that there is an important difference between the earlier examples and this new one, however; and this difference is supposed to legitimate their distinction. The difference is that this example involves a definite description where the others involved proper names.

This is a consistent Millian position, of course, but it encounters difficulties with the puzzle about belief. It may help to understand the Fregean position if it is seen as assimilating the cases involving names to the case of definite description. And it would be question-begging, it seems to me, to respond to my argument for this Fregean position by denying this alleged similarity between names and definite descriptions—whether there is such a difference is implicitly challenged by my argument.

To return to the main line of argument, I believe that Kripke successfully refutes the claim that certain standard examples by themselves demonstrate the unacceptability of (S). However, in order for that refutation to constitute an indirect defense of Millianism, as Kripke suggests, there must be no other principle involved in his own analogs which is, like (S), a consequence of Millianism. That, it seems to me, cannot be

---

19 It can seem more attractive in relation to a notion correlative to that of ambiguity, that is, the notion of *synonymy*. The idea would be that for names to be synonymous they need only share a reference while definite descriptions are synonymous only under logically stronger conditions: when their constituent terms are themselves coreferential, for example.
done. Something like (H) must be presupposed in order to obtain the contradictory results. Moreover, since (H), again like (S), is not a consequence of Fregeanism, the puzzles (indeed, Kripke’s puzzles now help in this regard!) constitute indirect support for a Fregean position.

The methodological situation has turned out to be as follows:

Someone wishes to give a *reductio ad absurdum* argument against a hypothesis (S). He seems to succeed in refuting the hypothesis, and although his derivation of an absurdity from the hypothesis may make use of auxiliary principles, such as (D), which he accepts as self-evident, the use of these principles is not essential in deriving the absurdity. Once we know that the same absurdity can be recreated without (S), we should not blame the contradiction on (S) (or on the auxiliary principles). If we notice, however, that there is one general principle, (M), such that the puzzle cannot be created without *some* consequence (e.g. (S) or (H)) of that principle (again, with or without (D) and (T)), then we have good reason to blame the contradictory results on (M).

I have put my point in terms of the hermeneutic principle and the steps that require it. This is one way of indicating my response to the puzzle.²⁰ Of course, there are other ways of putting what is essentially the same point. There is, for example, a way of transforming the response I am making here into a response that rejects the *applicability* of the principle of disquotation²¹ (whether or not the principle itself is rejected).

---

²⁰ It may be that step (7) of the traditional argument (step (9) of Kripke’s rendition), by contrast to steps (8) and (7) of the bilingual and monolingual analogs respectively, *would* satisfy an improved hermeneutic principle. If so, and if, as I have argued, the disquotation and translation principles are not essential in creating the puzzle, then the problem with the traditional argument *would* after all be its use of (S), *contra* Kripke. His examples are disanalogous, this view would proceed, in their using Fregean-ambiguous names where the traditional argument does not. The basic point of analogy, I claim, their ultimate reliance on (M), would remain. Thus, Kripke’s puzzle does not go far enough in its defense of (S). Ultimately the traditional argument (or a version of it) can be used (together with other considerations adduced here) to reject (S).

²¹ Here follows the discussion promised in footnote 16 above.
Remember that Kripke himself limits the applicability of that principle to cases in which the expression of the agent’s belief does not contain any ambiguities. When we apply and then reapply the disquotation principle for English to Peter and conclude that he believes that Paderewski has musical talent and that he believes that Paderewski does not have musical talent, is it not at least controversial whether the ordinary language term ‘Paderewski’ as Peter is using it has a single sense? And if it does have more than one sense, and if we are not presupposing Millianism, then is it not at least controversial whether the application of the disquotation principle is legitimate? (Again, this is not to reject the principle itself, merely to suggest that one of its presuppositions may not be satisfied in the relevant cases.)

Because I argue that the puzzle about belief is infected by a kind of ambiguity, one reaction to my resolution might be to attempt to recreate the puzzle while stipulating that there is no ambiguity (of any kind) in its expression. Of course, it is not plausible to suppose that Peter can rationally assent both to “Paderewski has musical talent” and to “Paderewski does not have musical talent” unless he attaches different senses to “Paderewski.” 22 But we can insist that we not admit any ambiguity in our representation of the puzzle. 23 That is, we can simply lay it down as a condition on our representation of the puzzle that none of the terms used in that representation be ambiguous. What are the consequences for the puzzle of this stipulation?

We can and do disquote people even when they use ambiguous expressions. However—and this is crucial—our use of an ambiguous expression in the belief

23 Kripke may be sensitive to this idea when he says (1988, p. 130) “…Peter will assent to ‘Paderewski had musical talent’, and we [emphasis his] can infer—using ‘Paderewski’, as we usually do [emphasis mine], to name the Polish musician and statesman…”
attribution should be disambiguated in a way that matches the original use of the ambiguous term by the person disquoted.24 Similarly we can translate sentences containing ambiguous expressions from other languages into sentences in our own language. However, if the sentence in our language is to preserve the truth value of the original, then the term we use to translate the ambiguous term should itself be disambiguated in a way that matches the disambiguation of the original term. “Banco” has a similar ambiguity in Spanish25 as “bank” does in English. If “el banco cierra a las cuatro” is true in Spanish even though the river banks never close, then, in order to guarantee that truth value is preserved, if we translate the sentence into English as “the bank closes at four,” we must disambiguate “bank” as something like “financial institution.”

This shows that we can allow a more liberal principle of disquotation—one that does not exclude ambiguity from the embedded clause. This version of the disquotation principle could be put as follows:

\[(D^\prime)\text{isquotation} \quad \text{If S, a normal L-speaker, on reflection, sincerely assents to ‘p’ (a sentence of L), and if ‘q’ is a translation (into English) of ‘p’ such that there is a sense of ‘q’ that is the same as the sense of ‘p’ in which the speaker assented to it, then S believes that q, in that sense of ‘q’}.\]

---

24 The text was going to read, “[H]owever, if our attribution is to be true, we must use the ambiguous expression in our belief attribution in a way that matches its use by the person disquoted.” This isn’t exactly right: we could get lucky. An anglophile zoologist says (orally), “I love whales.” He is reported (also orally) as loving Wales. He does—the attribution is true. But my written representation of the case shows that the oral attribution is to be disambiguated in a way that does not match the proper disambiguation of the sentence originally (orally) affirmed.

25 Actually, it’s many-ways ambiguous, meaning also bench, counter, school (as of fish), and referring to the game of Baccarat.
With this principle, we needn’t require, as Kripke has, that “the sentence replacing ‘p’ is to lack…ambiguities.”

Above I spoke of the need to “match” the sense of the expression embedded in our attribution to that with which the agent takes the sentence (to which he assents). In (D”) I interpret that relation of matching simply as identity: the sense of ‘q’ must be the same as that in which the agent assents to ‘p.’ The Millian, by contrast, will allow a translation of ‘p’ into any English sentence ‘r’ so long as ‘r’ satisfies the weaker condition that all of its parts have the same extension as the corresponding parts of ‘p.’ Millianism, in effect, interprets the relation of matching as that of co-referentiality of parts. This raises an interesting question: is there a continuum between the identity of the senses of the parts of two sentences, at one end, and the co-referentiality of the senses of the parts of those two sentences at the other? Here I will limit my discussion to the most extreme Fregean position that requires identity of sense for proper disquotation.

When Pierre assents to “Londres est jolie” and we disquote him in French and conclude “Pierre croit que Londres est jolie,” the sense of our “Londres” must be identical to Pierre’s. Similarly, our translation of “Pierre croit que Londres est jolie” into “Pierre believes that London is pretty” must preserve this match so that our use of “London” now matches his use of “Londres.” But when Pierre assents to “London is pretty” it should be clear that he takes “London” there with a different sense than that with which he took “Londres” earlier. To disquote him properly we would have to use this new sense in attributing to him the belief that London is pretty. But since we have stipulated that we will use terms unambiguously, we are blocked from making that attribution—at least, we are blocked from making it with the term ‘London.’ Likewise if Peter did not attach different senses to “Paderewski,” he would not assent to both “Paderewski has musical talent” and to “Paderewski does not have musical talent.” But if he does attach different senses to that word, then to disquote him properly, we must
preserve that ambiguity in a corresponding way in our belief attributions. Our earlier stipulation would block us from preserving that ambiguity (or would prohibit us from using the same name ‘Padarewski’ on both occasions).

At one point (p. 131) Kripke notes, “If Peter had not had the past history of learning the name ‘Paderewski’ in another way, we certainly would judge him to be using ‘Paderewski’ in a normal way, with the normal reference, and we would infer [his belief] by the disquotation principle” [emphasis mine]. It does seem plausible that the earlier and later history of the subjects in these examples should not be relevant to our practices of disquotation. But now we have a principled reason for letting those earlier and later histories affect our belief attributions. Once it is established that he is not using ‘Paderewski’ in a normal way, with the normal reference (Peter uses it as if it were referentially ambiguous when it is not), we cannot use ‘Paderewski’ as we normally do to report Peter’s beliefs. Peter has not changed his mind about anything he believed before, and if a shock wiped out all of his memories he would be like anyone else (whom we would disquote normally) that assents to “Paderewski had no musical talent”; but the changes do make a difference with respect to the propriety of using certain words as we normally do in attributing beliefs to him. In sum, we might say that the puzzle about belief is the problem of the essential (Fregean) ambiguity.

A challenge and reply

Here it is important to respond, if only briefly, to a challenge Kripke raises for any resolution of the puzzle that appeals essentially to the theoretical construct of senses. Kripke is sceptical of any such account. According to him, “the clearest objection [to a

Fregean view], which shows that the others should be given their proper weight, is this: the view under consideration does not in fact account for the phenomena it seeks to explain” (p. 110). I will not try to argue that there is something which should not be given its proper weight (and we don’t need a clear objection to show that anything should be given its proper weight). But I will argue that the Fregean view does account for the phenomena it seeks to explain.

There is widespread sentiment that Kripke (together with Donnellan and others) has decisively refuted a traditional view of names.27 I too am completely convinced by Kripke’s arguments. They decisively refute the position against which they are directed. That position, however, is not Frege’s. Indeed, it seems to me that what is often called ‘the Fregean position’ is one Frege was at some pains to disown.

Compare three representations of the Fregean theory with its original presentation.

The … Fregean view holds that to each proper name, a speaker of the language associates some property (or conjunction of properties) which determines its referent as the unique thing fulfilling the associated property (or properties). This property(ies) constitutes the ‘sense’ of the name. … Frege (and Russell) concluded that, strictly speaking, different speakers of English (or German!) ordinarily use a name such as ‘Aristotle’ in different senses (though with the same reference). (Kripke, 1988, p. 103)

The most natural model for the sense of a proper name is one in which it is a concept the user associates with the name. … [G]iven the fact that our concepts tend to differ, … it will be a common occurrence that the senses we associate with our words differ …. (Richard, pp. 64, 65)

Frege is identifying the information value of the name for a particular user with the purely conceptual content the user associates with the name. … [T]he conceptual content attached to a proper name, as used with a particular reference, varies significantly from speaker to speaker. (Salmon, 1986, pp. 47, 125)

[T]here [is] connected with a sign (name, combination of words, letter[!]), … what I should like to call the sense of the sign, wherein the mode of presentation is contained. … The same sense is not always connected, even in the same man, with the same idea. The idea is subjective: one man’s idea is not that of another. … This constitutes an essential distinction between the idea and the sign’s sense, which may be the common property of many and therefore is not a part or a mode of the individual mind. … In light of this, one need have no scruples in speaking simply of the sense …. (Frege, 1960, pp. 57, 59–60)

We must appreciate Frege’s insistence that senses are objective.

Many appeal to Frege’s second footnote to “On Sense and Reference” to support the interpretation that the sense of a name can vary intersubjectively. But Frege is careful: “In the case of an actual proper name such as ‘Aristotle’ opinions as to the sense may differ. It might, for instance, be taken to be the following: the pupil of Plato and teacher of Alexander the Great. Anybody who does this will attach another sense to the sentence ‘Aristotle was born in Stagira’ than will a man who takes as the sense of the name: the teacher of Alexander the Great who was born in Stagira” (1960, p. 58n*, emphasis mine).

---

28 See, e.g. Kripke (1988, pp. 136n2, 137n3) and Richard (p. 64).
29 Unfortunately, in that same footnote, Frege is also incautious, describing the ‘Aristotle’ case as an example of “variations of sense.” Frege (1988, pp. 40–2) also seems to me to support my interpretation.
This all suggests not so much that the sense of ‘Aristotle’ varies intersubjectively, but that opinions as to the sense, or what is taken as the sense, can vary intersubjectively. Remember, “the same sense is not always connected, even in the same man, with the same idea.” A single sense can be connected with different ideas. With this in mind, let us reconsider Kripke’s objections.30

The objections directed against what Kripke calls “the extreme Frege-Russellian view that the senses of proper names vary, strictly speaking, from speaker to speaker” will not concern us. As I have said, that extreme view is one, it seems to me, Frege was careful to distinguish from his own. Frege in effect agrees with Kripke that “[n]ames…are the common currency of our common language” (Kripke, p. 111) when he says that “the sign’s sense…may be the common property of many” (Frege, 1960, p. 59).

For Frege, the sense of a name determines its referent. Perhaps the most plausible Fregean ‘description theory’ of names is one according to which names are associated with uniquely identifying properties. Later Kripke considers this view. “It is noteworthy,” he claims, “that the puzzle can still arise even if Pierre associates to ‘Londres’ and ‘London’ exactly the same uniquely identifying properties” (p. 125). We are invited to imagine a case in which Pierre associates with ‘Londres’ and with ‘London’ the properties being the largest city in (and capital of) England, containing Buckingham Palace, etc. For our purposes, what is important is that there be no property constituting the sense of one of the names that does not also constitute the sense of the other. And Pierre is supposed to grasp this one sense and associate it with both ‘Londres’ and ‘London.’ In such a case neither ‘London’ nor ‘Londres’ would be ambiguous, even by Fregean standards.

30 Recall my p. 37n22.
Now, according to Kripke, Pierre “need only conclude that ‘England’ and ‘Angleterre’ name two different countries, that ‘Buckingham Palace’ and ‘le Palais de Buckingham’... name two different palaces, and so on.” He concludes (p. 126), “[t]he fact is that the paradox reproduces itself on the level of the ‘uniquely identifying properties’ that description theorists have regarded as ‘defining’ proper names (and a fortiori, as fixing their references).”

I am sceptical of Kripke’s conclusion. Is it possible for Pierre to be associating the same property with both ‘London’ and ‘Londres’ if he (so to speak) believes Angleterre and England to be different countries and if the property he associates with ‘Londres’ he would express using ‘Angleterre’ (and not ‘England’) while the property he associates with ‘London’ he would express using ‘England’ (and not ‘Angleterre’)? I don’t think so. And it may be that an affirmative answer prejudges the central question at issue.

To clarify the dialectic: I am considering the prospects for a Fregean position concerning the semantic value of proper names. Kripke claims that the position does not really resolve the puzzle since the puzzle rearises for that position. But does it? If a single uniquely identifying property can be the sense of two names α and β, and if Pierre can apparently without contradiction affirm $\cup F_{\alpha'}$ and $\cup \sim F_{\beta'}$, then the puzzle rearises. 31

Pierre can apparently affirm “London is not pretty” and “Londres est jolie.” Can a single uniquely identifying property be the sense of ‘London’ and ‘Londres’ in his affirmations? Kripke would argue that it could, that the sense could be composed of properties such as being the capital of England, etc.

31 It is significant that it seems impossible to create a ‘monolingual’ version of this puzzle on the model of the Paderewski puzzle. Such a puzzle would have a single name, $\alpha$, with a single sense such that Pierre can apparently without contradiction affirm $\cup F_{\alpha'}$ and $\cup \sim F_{\alpha'}$. Kripke does not offer such a puzzle.
My response is that ‘being the capital of England’ has an unambiguous meaning only if ‘England’ has an unambiguous meaning. In the example, of course, ‘England’ refers unambiguously. But by now we know better than to assume that unambiguous reference implies unambiguous meaning: such an assumption seems to presuppose Millianism. According to the view I defend, in this new example Pierre is associating two different properties with (respectively) ‘England’ and ‘Angleterre,’ both of which can be expressed by the (consequently ambiguous) English expression ‘England.’ That this is Pierre’s situation is confirmed when he fails to draw ordinary logical consequences of what would be the conjoint set of his (occurrent, reflective) beliefs if only he did associate the same property with both ‘England and ‘Angleterre.’

If he really associated the very same uniquely individuating property with ‘England’ and ‘Angleterre,’ as he must if he is to associate the same property with ‘London’ and ‘Londres,’ then he could not “conclude that ‘England’ and ‘Angleterre’ name two different countries…” as Kripke suggests. Without some distinction between the individuating properties associated with the two names ‘London’ and ‘Londres,’ Pierre would be in no position to draw the conclusion Kripke suggests: how could he think, for example, that ‘London’ is a name for the capital of England and that ‘Londres’ is a name for the capital of England and not think that ‘London’ and ‘Londres’ name the same city (viz., the capital of England)? How could he think that ‘England’ is a name for the F and that ‘Angleterre’ is a name for the the F and not think that ‘England’ and ‘Angleterre’ name the same thing (viz., the F)?

So the view I defend can reject the attempted rehabilitation of the puzzle at the level of uniquely identifying descriptions. It is a testament to the depth of Kripke’s article that he goes on to consider the position that underlies this rejection.

Of course the description theorist could hope to eliminate the problem by ‘defining’ ‘Angleterre’, ‘England’, and so on by appropriate descriptions also. Since in principle the problem
may rear its head at the next ‘level’ and at each subsequent level, the description theorist would have to believe that an ‘ultimate level’ can eventually be reached where the defining properties are ‘pure’ properties not involving proper names (nor natural kind terms or related terms…). I know of no convincing reason to suppose that such a level can be reached in any plausible way, or that the properties can continue to be uniquely identifying if one attempts to eliminate all names and related devices. (p. 127)

The footnote to this passage notes that “[t]he ‘elimination’ would be most plausible if we believed, according to a Russellian epistemology, that all my language, when written in unabbreviated notation, refers to constituents with which I am ‘acquainted’ in Russell’s sense” (p. 143n29).

I agree with Kripke that we must appreciate the connection between (i) (what Kripke calls) the description theory, (ii) the acceptance of a level of ‘pure’ properties, and (iii) (what he calls) a Russellian epistemology. It seems to me that there are some considerations that may make us less pessimistic than Kripke about the prospects for such a combination of views.

First, there may be a category error in the claim that the description theorist must believe a level of pure properties not involving proper names can eventually be reached. It is not clear how any property could involve a proper name.\footnote{Unless it were the name type figuring in the property; but then it would be name tokens that would satisfy the property and not ordinary individuals.} Prima facie, only expressions can involve names. Suppose then that what is meant is that the pure properties will be expressed by terms that do not involve any proper names.

Now, if Kripke is right that the puzzle can in principle rear its head at each subsequent level, then he is right that the description theorist must postulate a level of pure properties whose expression will not involve proper names. But according to a
Russellian epistemology, there are ‘logically proper names’ with whose referents we are ‘acquainted.’ If the puzzle cannot rear its head at the level of logically proper names, then the level of pure properties can include properties that would be expressed by logically proper names. Russell thought that we were acquainted with universals. So it would be sufficient for my position that a level of universals could be reached.

Second, Kripke doubts that properties can continue to be uniquely identifying if names and related devices do not figure in the terms in which they are expressed. If that were the case, sense would not determine reference and the view I defend would be refuted. On the other hand, if the metaphysical principle of the identity of indiscernibles is true, then it is impossible to produce a case of two individuals that shared a single set of properties. The general possibility of purely qualitative uniquely identifying descriptions would be established.33

Third, privacy is not a necessary concomitant of acquaintance. Whatever Russell may have claimed, it is clear that two agents can be acquainted with the very same property or universal (these must not be confused with sense-data). Hence, the position I defend does not have the consequence that “no one speaks a language intelligible to anyone else” or that “no one speaks the same language twice,” as Kripke fears (p. 143n29).

Intermediate conclusions

I do not claim to have established the Fregean view of the semantic content of names in belief contexts. I hope, however, to have made it more attractive and to have cleared up

33 This connects with the underappreciated fact that traditional Fregeans must find a way to reject or reinterpret Black’s example of qualitatively identical (but numerically distinct) spheres. See Black (1960).
some possible misunderstandings. Kripke’s puzzle seems to presuppose Millianism. If
my argument is correct, examples of apparent failures of codesignative names to be
interchangeable in belief contexts, as in the traditional argument, can be taken to argue
against the Millian view. At least, that view seems to be inadequate as it stands. Perhaps
it could be supplemented without being rejected.

Moreover, the opposing Fregean view accommodates those examples. Many of
Kripke’s negative arguments in “A Puzzle About Belief” and in Naming and Necessity
are directed against the position that the senses of names vary from speaker to speaker.
This is neither an attractive, nor, it seems to me, Frege’s position. Better is a view that
senses are “the common property of many,” that senses are objective features of the
names with which they are associated, and that the sense of a name determines its
reference.

Accordingly the view would hold that the sense of a name is a property that
uniquely identifies an entity as the reference of that name. I have responded to Kripke’s
argument that the puzzle rearises for this view. And it seems to me that the prospects for
developing the view, along lines suggested by Kripke himself, may be less dim than he
and others suppose. A promising line of development would combine the Fregean
semantics sketched here with a Russellian (or Fregean or Gödelian) epistemology of
‘acquaintance’ (or ‘grasping’ or ‘intuiting’) and a metaphysics that underwrites Leibniz’s
principle of the identity of indiscernibles. Such a development remains to be seen; but so
does its refutation.

Earlier I said that I would (i) locate clearly where the puzzle about belief lies and
in what it consists, (ii) carefully consider the methodology of Kripke’s paper, and use that
methodology for my own purposes, and (iii) investigate the relations between the
concepts of rationality, meaning, and reference. What are the results?
(i) The puzzle is located in the tension between apparently innocent assumptions and the contradiction to which these assumptions seem to lead. Much of the early discussion was devoted to showing exactly which assumptions are needed and how the derivation proceeds. My presentation of the traditional argument and of the analogs does deviate from Kripke’s own narrative (I do not present it as involving a choice among four exclusive and exhaustive but unattractive options)—but in a way which seems to me only to bring out more clearly the way in which his use of the puzzle exemplifies the methodology he takes himself to be exploiting.

(ii) Kripke’s methodology is very much like that of a scientist. In effect his paper makes the following accusation: in the traditional argument the anti-Millian uses the results of a thought experiment to reject Millianism without properly controlling for the possible effects of other factors. This is an acceptable methodology. And it is true that when we perform thought experiments that control for the effect of (S), we obtain results similar to those obtained in the traditional argument. Moreover, and as against Kripke, we can control for the effects of (D) as well and still obtain similar results. Kripke’s own Paderewski example controls for the effects of (T). This suggests that the results are due neither to (S) nor to (D) or (T). I try to show that when we do control for the effect of (S), however, we must exploit a hermeneutic principle (H), concerning the logical representation of ordinary language names, that presupposes Millianism (but that could be rejected by Fregean theories). The thought experiments demonstrate that some Millian principle or other is required to produce the results and this shows that the puzzle about belief derives from Millianism. Thus,

(iii) If we limit the semantic value of a name to its reference, as in Millianism, we will encounter difficulties with respect to the concept of rationality: either we will have to say that it is not irrational to have contradictory beliefs (even when the beliefs are fully occurrent and reflective) or we will find ourselves characterizing as irrational people who
are clearly not irrational. By contrast, if we identify the meaning of a name with its sense, Frege’s theoretical innovation, we are not puzzled about belief.

The moral is that Millianism has substantial consequences in addition to its implications about substitutivity. Shakespeareanism is only one symptom of the syndrome afflictng Millianism. The position also makes nonsense of a natural and attractive use of logical notions in epistemic judgments. If we want it to be a condition on rationality that an agent not contradict himself, or at least not do so fully occurrently and reflectively, in other words, if we want “anyone, leading logician or no, [to be] in principle in a position to notice and correct contradictory beliefs if he has them” (Kripke, p. 122), then Millianism cannot be the whole story.

At this point, it may seem to you that in resolving one difficulty, I have forced us into a deeper difficulty. I have urged that a Fregean theory of meaning is needed to account for some powerful intuitions about rationality and the semantics of belief ascription. Nevertheless, it is well known that since the early 1970’s Fregean theories have been subject to a sustained and serious critique that can hardly be ignored. Now if, contra Kripke, Millianism is differentially subject to a traditional difficulty, then we seem to be at an impasse. Both of the leading theories of meaning have some strengths; but each seems to have unacceptable weaknesses. I would like to go on now to indicate the direction in which I believe progress can be made.

Part II

A hypothesis

Hume said, “from this circumstance alone, that a controversy has been long kept on foot, and remains still undecided, we may presume, that there is some ambiguity in the expression and that the disputants affix different ideas to the terms employed in the
controversy” (Enquiry, §VIII). My suggestion is simply an application of Hume’s insight. If some issues in the philosophy of mind and language seem to have stalled, it may be because there is some ambiguity in the expression of the controversies and the disputants affix different ideas to the terms employed.

Begin with a supposition. What if two notions that are combined in the traditional understanding of proposition were separated. Traditionally, a proposition is what we have in mind in having a so-called propositional attitude as well as what we say in uttering a declarative sentence. Propositions are taken to play both roles at once. This idea is sustained by the platitude that what we say, when we assert a declarative sentence (assuming that we are competent, sincere, have reflected, etc.), is precisely what we believe; that is, the semantic value of the sentence we assert is supposed to be identical to the entity to which we are related by belief. I propose to reject that picture.

Rejecting this picture introduces an ambiguity into the notion of proposition. In one sense, what it is to be a proposition is to be the object of a cognitive attitude. In this sense, propositions are most basically what we have in mind when we have a thought. In another, independent sense, what it is to be a proposition is to be what is said by a declarative sentence. In this other sense, a proposition is most basically the piece of information asserted by the writing or uttering of a declarative sentence.

Now, suppose that corresponding to that deep ambiguity in the notion of proposition there were two semantic theories. Suppose that according to one theory, the meaning of a proper name, in the sense of the name’s contribution to propositions expressed by sentences in which it occurs, is just its referent. That theory would hold that a proposition can be constituted inter alia by the referent of a proper name used in a sentence that expresses it. A theory that held, on the other hand, that the meaning of a proper name is not its referent but something else, a sense, would hold that propositions
are constituted not by the referents of the proper names used in their expression, but by
the senses of those proper names. What I am asking you to consider, of course, is that
Millianism and Fregeanism are the theories that correspond, respectively, to the intuitions
concerning what we say and what we have in mind.

Notice that by separating these notions as I would, the following thesis is rendered false (let $L$ range over languages and let ‘$p$’ be a substitutional variable for sentences of $L$): a sincere, reflective subject who is competent in the use of $L$ and who affirms $\cup p'$ believes that $p$. How could it be that this thesis is false? This will be clarified in greater detail by the discussion to follow. But to make the point briefly, too much weight is carried by the notion of competence. What it is, exactly, to be a competent user of a language is obscure to me. But one thing I’m sure of: it does not require that one know the referent of every proper name in the language. When we use language, we immerse ourselves in a public system. This system is governed by convention. These conventions can outrun, it seems to me, almost any degree of competence in the use of the language. As a result, it is possible to ‘misspeak’ even when we are competent and sincere and reflective. A sincere, reflective subject who is competent in the use of $L$ might affirm $\cup p'$ thinking, mistakenly, that it represents her belief that $p$. And her being wrong about whether $\cup p'$ represents her belief that $p$ does not by itself undermine her competence in using $L$.

What would be the consequences of the hypothesis that there is a basic difference between what is said by a declarative sentence and the object of a cognitive attitude and that Millianism and Fregeanism are, respectively, the theories that correspond to the intuitions concerning those different notions? The most significant consequence is that Millianism and Fregeanism would no longer be incompatible theories of the same phenomena but actually independent theories of distinct phenomena. Their apparent opposition would be a matter of wholesale intertheoretic equivocation.
Strengthening the hypothesis

The general hypothesis I am considering would be strengthened by connecting the two theories with two basic and abstract ideas or perspectives on the very goal or nature of a theory of meaning. What is the aim of a semantic theory? It may be a platitude that language serves the dual role of representing ways the world is and of communicating our thoughts by expressing them. The idea that language would serve such a dual role is made more coherent by the idea that our thoughts themselves represent ways the world is. So the aim of a theory of meaning might be to explain how it is that language can both represent the world and express our thoughts.

But language *could* serve one role without serving the other. That is, language could represent ways the world is without expressing our thoughts. This could be so, for example, if our thoughts represented the world *indirectly*. If our thoughts bore the representation relation to ways the world is by bearing a different relation to some other things, properties of ways the world is, for example, then there would be the possibility that we could have two different thoughts that represented the same way the world is, that is, we could have thoughts bearing relations to two *different* properties of the *same* way the world is.

If language were to serve only the purpose of representing ways the world is, there would be no point in having different bits of syntax for those two thoughts. And our semantics could abstract away from any difference in syntax that actually existed in a given natural language. To the extent that what is really important in *communication* is to transmit a certain kind of *information*, that is, to inform another thinker as to how the world is, to that extent, the purpose of language is to represent ways the world is.
I say this by way of suggesting what may be the basic idea underlying a Millian semantics. To the extent that language developed, and it may be helpful to think of this in evolutionary terms, in order to allow one person to inform another where the water is, to let one person tell another that this animal is dangerous and that one is edible, and so on, to that extent language can be seen as nothing more than a vehicle for the expression of singular propositions. To the extent, however, that language is meant to express the contents of our thoughts, and if we could have two different thoughts that nevertheless represent the same part of the world as being the same way, then we are pulled toward a finer-grained individuation of propositions.

A basic idea of Millian semantics is that the difference between expressions such as “Hesperus is bright” and “Phosphorus is bright” that represent the same part of the world as being the same way is not a semantic difference. If the purpose of language were to represent states of the physical world, it could be served equally if there had not been those two different names. And we could abstract away from the different syntax in our semantic theory. But the Fregean semanticist is more impressed by the role of language in representing the state of mind of thinkers; Fregeanism emphasizes an agent’s use of language in expressing her thought, in “speaking her mind.”

According to the Fregean, to the extent that two expressions can represent different states of mind that nevertheless (in turn) represent the same part of the world as being the same way, to that extent the two expressions are different in a way that should be captured by the theory of meaning. After all, the very point of a theory of meaning, according to this perspective, is to explain the way language expresses our mental states. And this may be the perspective that underlies Fregeanism. That the expressions “Hesperus is bright” and “Phosphorus is bright” correspond to different contents is sustained by the intuition that it would not be epistemically censurable (at least not in the
most serious way) to believe what is expressed by one while disbelieving what is expressed by the other.

To put it another way, two people who are related by belief to all and only the same Russellian propositions might still be different in an important way—one might be rational and the other irrational. One might be open to censure in a way that the other is not. Here we rely on intuitions about the agent’s mental or epistemic state. And this, it seems to me, is to rely on a set of intuitions of a very different kind from those that sustain Millianism. For although one might be rational while the other is not, the truth conditions of their respective beliefs will be the same.

Here I would propose some new terminology. No longer will we use any of the controversial terminology: “meaning,” “belief,” “content,” “proposition,” “ambiguity,” “synonymy,” “validity,” and “contradiction” are all banned. For each of those terms we introduce two terms: for “meaning” we introduce “F-meaning” and “M-meaning,” for “belief” we introduce “F-belief” and “M-belief,” and so on, corresponding to the understanding of that term in the Fregean theory and Millian theory respectively.

Now, as a foil, consider the artificial notion of \(t\)-content. Imagine a theorist who held that the semantic value of a declarative sentence is nothing more than its truth value—the proposition expressed is just the truth value. Suppose that two expressions have the same \(t\)-content if and only if they have the same truth value. Of course, \(t\)-content is only a partial function of expressions—only grammatical sentences will have any truth value at all. We can, nevertheless, artificially define the relation of \(t\)-belief on the basis of \(t\)-content. A subject \(t\)-believes that \(p\) if and only if he bears the belief relation to the \(t\)-

---

34 You may notice an analogy between (i) the difference between Millianism and Fregeanism with respect to names (what the former takes as the only semantic value, the latter takes as a secondary semantic value, to be determined by another) and (ii) the difference between this foil and Millianism.
content of the embedded sentence (that is, if and only if he believes the embedded sentence to be true). It follows that simply by believing some true sentence, an agent will t-believe that p for any true sentence \( \varphi p' \). Suppose that (i) S t-believes that p; then any sentence \( \varphi q' \) with the same truth value as \( \varphi p' \) could be substituted for \( \varphi p' \) in (i) _salva veritate_.

I introduce the artificial notions only to make a point: there is _some disadvantage_ in having t-contradictory t-beliefs—it entails error. Unless all of your t-beliefs have the same t-content, then at least one of your t-beliefs is false.\(^{35}\) If you have beliefs whose contents vary in truth values, then you have beliefs not all of whose contents are true. But, clearly, the kind of epistemic defect exhibited by agents with t-contradictory beliefs is not so grievous. No doubt we all are at fault. I suggest that there is a continuum from t-contradictory t-belief, through M-contradictory M-belief, to F-contradictory F-belief.

Instead of trying to judge what may be an incommensurable set of demands on the notion of proposition, I propose we consider the possibility that mental representation and linguistic representation are two very different _modes_ of representation. Propositions in the sense of _what we have in mind_, the notion which is sustained by intuitions about rationality, are mental representations. These propositions are constituted exclusively by Fregean senses and represent the physical world _indirectly_, in virtue of being satisfied by some states of the world (or by some worlds) and not by others. Propositions in the sense of _what we say_, the notion which is sustained by intuitions about communication and information, are linguistic representations. These propositions abstract away from differences in Fregean propositions that are not correlated with differences in ways the

---

\(^{35}\) An important difference between t-contradictory beliefs and M-contradictory beliefs is that if you have t-contradictory beliefs, it might still be true that your beliefs _could_ have all been true. M-contradictory beliefs are not such that they could have been true together.
world is. They represent the physical world *directly*. Deep conflation of linguistic representation and mental representation may be at the root of lasting difficulties in the philosophy of mind and language.

Reinventing a relation

Notice the relation between Millianism and Fregeanism under the conception I have been recommending. They are independent theories that use, unfortunately, some of the same theoretical terminology. Of course each theory could simply introduce new terminology for the concepts of the other theory. So Millianism, whose conception of simple propositions takes them to be so-called Russellian propositions (containing an ordinary object ‘in subject place,’ so to speak), could simply use the expression ‘Fregean proposition’ for the conception of proposition that appears in Fregean theories.

Now, Millianism has a conception of meaning such that the meaning of a proper name is simply its referent. Fregeanism of course simply calls that the ‘referent’ of the proper name and identifies the meaning of the name with its sense. Based on *its* conception of meaning, Millianism defines the notions of ambiguity and synonymy. The theory develops to include a semantics of belief ascriptions. Assuming the thesis of *relationality*, belief ascriptions report relations between a subject and an entity constituted *inter alia* by the referent of a proper name.

Fregeanism develops independently. Since the Fregean’s conception of meaning is such that the meaning of a proper name is not simply its referent but something else, a sense, the definitions of ambiguity and synonymy will of course be very different—as we

---

36 I use “meaning” and “semantic value” neutrally (and interchangeably) to mean *what is contributed to the content of more complex expressions* (e.g. sentences).
have seen. Fregeanism produces its own theory of belief attribution. Belief ascriptions report relations between a subject and an entity that can be constituted not by the referent of the proper name used in the ascription but by the sense of that proper name.

Recall that because the theories have different understandings of what it is for a term to be ambiguous, the theories will differ with respect to when it is acceptable to represent a term by a single logical constant. Ordinary language terms which are unambiguous according to Millianism and thus apt for representation by a single logical constant may be ambiguous according to Fregeanism. Thus, arguments in ordinary language which satisfy the presuppositions of validity laid down by Millianism will not necessarily satisfy those laid down by Fregeanism. Still those arguments that do satisfy the presuppositions of validity of Millianism have an interesting and desirable property that it is not beyond a Fregean’s theory’s ability to appreciate: it can call those arguments “extensionally valid.” For example, it is extensionally valid to move from “Cicero is bald” and “Tully is bright” together with “Cicero=Tully” to “Cicero is bald and bright.”

If this were the situation, how could we choose between Fregeanism and Millianism? Briefly, it is not clear to me that there is here anything to choose between. But if we could compare them, it would be, I think, by appealing to a connection between semantic theory and epistemology—a connection that some semanticists would reject. The connection I have in mind is between semantic value and rationality. As argued earlier, Millianism encounters difficulties with the notion of rationality. So perhaps each theory is better suited to some purposes and worse for others.

Applications to traditional puzzles

What insight do we gain, with my proposal, into the semantic puzzles? What to say about for example Frege’s puzzle? Perhaps the best thing to say is that there is no one Frege’s
puzzle: there is an apparent puzzle created by the conflation of two non-puzzles. Call the two non-puzzles the non-puzzle about F-content and the non-puzzle about M-content. The non-puzzle about F-content is how it can be that two expressions with the same F-meaning can have different F-contents. But no two expressions with the same F-meanings can have different F-contents. “Hesperus is Phosphorus” and “Phosphorus is Phosphorus” do not have the same F-meanings, so the fact that one can be F-informative while the other is not poses no problem for the theory.

Similarly, the non-puzzle about M-content is how it can be that two expressions with the same M-meaning can have different M-contents. But no two expressions with the same M-meanings can have different M-contents so there’s no problem here either. “Hesperus is Phosphorus” and “Phosphorus is Phosphorus” have the same M-meanings alright; but they have the same M-contents too. Neither is more M-informative than the other.

Now, it is true that two expressions with the same M-meanings can have different F-contents; but why should that be surprising? On the contrary, when we see what it is to be an M-meaning or an F-content, it will be clear that there will be many such cases. Sentences that have the same M-contents can be differentially F-informative.

While the difference between M-meaning and F-meaning and some of the other new semantic terms may be relatively clear, the difference between F-informativeness and M-informativeness may be less so. Here’s a way of understanding that difference: a sentence is M-informative for a subject iff it expresses an M-proposition which is not already M-known by that subject. Correspondingly, a sentence is F-informative for that subject iff it expresses an F-proposition which is not already F-known by that subject. It should be clear how a sentence could be F-informative without being M-informative. It’s easier to be F-informative because M-propositions are less finely individuated.
The hypothesis I mean to be entertaining is that there are really two independent semantic theories each of which imperfectly accommodates almost all of the relevant intuitive data. Indeed, since the theories are independent they could in principle simply be conjoined, with the appropriate terminological legislation. Such a strengthened theory, it seems to me, could accommodate almost all of the relevant intuitive data more perfectly. Of course there is one aspect of the intuitive situation that is ill-accommodated by such a strengthened theory. That is the intuition that there is a single type of thing that best deserves to be called the “meaning” of a sentence, a single type of thing that deserves to be called the “proposition” expressed by a sentence, and so on—that our pre-philosophical uses of “meaning,” “belief,” etc. are univocal. But how strong is that intuition? Indeed is that really the type of thing about which we can legitimately be said to have intuitions? Isn’t that a theoretical point?

The idea that there are really two independent theories imperfectly upheld by the relevant intuitions may become still more appealing when one is faced with the kind of response Millians give to a characteristic criticism. On the Millian view, if Jones believes that Mark Twain is a wonderful author, she believes that Samuel Clemens is a wonderful author. That is an extremely counterintuitive result of Millianism. One way Millians try to limit the force of that counterintuitive result is in terms of what they call the “BEL” relation. The BEL relation is a three-place relation holding between thinkers, M-propositions, and ways of taking those M-propositions.

How does the introduction of the BEL relation limit the force of the counterintuitive result? Simply because the Millian can now allow that there are many modes of presentation of the M-proposition in question which do not figure in any BEL relation in which Jones is involved. For example, assuming for the sake of argument that

---

37 Salmon (1986).
linguistic expressions themselves can be modes of presentation, the following relation does not hold: (i) BEL (Jones, Clemens is a wonderful author, “Clemens is a wonderful author”). Only this relation holds: (ii) BEL (Jones, Clemens is a wonderful author, “Twain is a wonderful author”). But since the belief relation is postulated to be the existential generalization (on the third term) of the BEL relation, (ii) suffices in the Millian theory for Jones to believe that Clemens is a wonderful author.

Modes of presentation of M-propositions could for present purposes be identified with F-propositions. 38 Thus some Millians introduce theoretical materials that suffice for the production of Fregeanism. A diplomatic Fregean could simply create a “BEL*” relation based on the BEL relation: this BEL* relation would be a two-term relation holding between thinkers and modes of presentation of M-propositions, i.e. the kinds of the things that are in the domain for the first and third terms of the BEL relation. 39 Now, the BEL* relation is just F-belief.

This idea can be pressed into service to deepen my response to Kripke’s puzzle. I have claimed that even when Pierre believes that London is pretty and that London is not pretty, he may not have contradictory beliefs. And this may be true even when “London”

---

38 I assume, what may be controversial, that a mode of presentation of an M-proposition determines a unique M-proposition. I do not assume that the third term of the BEL relation will determine the M-proposition that is its second term.

39 The third term of the BEL relation is not meant to determine the second. If it were, then the BEL relation would be redundant anyway and could (should for theoretical elegance) be replaced by the BEL* relation. That the third term is meant to determine the second would be suggested by calling that term a ‘mode of presentation’ of the Russellian proposition which is the second term of the relation; but it is usually called a ‘way of taking’ the M-proposition. Modes of presentation, at least in Frege’s use of that term, are supposed to determine that which they present. On the other hand, insofar as Millians are attracted to a causal theory of names, the third term of the BEL relation will not fully determine the second term: that second term will be partly determined by causal relations involving the thinker (the first term), causal relations that hold independently of the third term. I discuss this and related issues more thoroughly in Chapter 2, pp. 93–8.
is unambiguous in the sense of not having more than a single definite reference. Now these claims can be refined and placed in a broader theoretical context. What I could now say is that Pierre can F-believe that London is pretty and that London is not pretty and not F-contradict himself, even when “London” is not M-ambiguous. Of course, if “London” is not M-ambiguous, “London is pretty” and “London is not pretty” will be M-contradictory and Pierre will M-contradict himself. This should recover any residual feeling that there is something contradictory about Pierre’s epistemic state. But so long as “London” is F-ambiguous, and it does not seem to me farfetched to suppose that it is in the example as given, the F-contents of Pierre’s F-beliefs are not F-contradictory.

This bifurcation of semantic theory may infect epistemic concepts. If whether or not you are rational is a function of your beliefs, and in particular is partly a function of relations among the contents of your beliefs, then if there is more than one kind of content, there is correspondingly more than one kind of rationality. It is easy to imagine a distinction between F-rationality and M-rationality. It suffices for M-irrationality, for example, that the M-contents of a pair of your M-beliefs be M-contradictory; for F-irrationality, on the other hand, the F-contents of a pair of your F-beliefs must be F-contradictory.

If the equivocation extends into epistemology in this way, then the argument against Millianism developed earlier admits of a subtle reply. The traditional argument and Kripke’s bilingual and monolingual analogs turn out to be equivocations on the term “rational.” What the examples all show, on this interpretation, is that someone can be F-rational without being M-rational. If M-rationality is a kind of rationality that deserves the name, then the earlier examples do not after all constitute indirect evidence for Fregeanism over Millianism—they show only that Millianism must be supplemented by something like a Fregean theory (as it often, in effect, already is—by way of the BEL
relation). No longer should we say that Millianism encounters difficulties with the notion of rationality. The point is rather that it deploys a different concept of rationality.

Millianism, on this approach, no longer forces us to say that it can be rational to have contradictory beliefs; nor does it force us to characterize as irrational people who are obviously not irrational. Rather, on this approach, Millianism claims that it can be F-rational to have M-contradictory beliefs; and it characterizes as M-irrational some people who are obviously not F-irrational. But Fregeanism should agree with that claim and with the characterizations.

I doubt, however, that M-rationality is any kind of rationality. Notice that, using our earlier notions of t-belief and t-contradiction, we could define a notion of t-irrationality which required only that a pair of t-beliefs be t-contradictory. But t-irrationality is no kind of irrationality at all: it is simply error. M-irrationality is worse than t-irrationality of course: it requires you to believe one and a contrary thing about a single entity. Your beliefs are t-contradictory if you believe that DNA is a double helix and you believe that Argentina won the World Cup in 1990. But those beliefs are not M-contradictory. For that you must believe of, for example, a certain person, that he is and is not a pillar of the community.

Still, not even M-contradiction should imply any serious epistemic censure. Although Millians deny that a sentence with a definite description as subject expresses a Millian proposition, there is (perhaps relative to context of utterance) a Millian proposition associated with whatever proposition is expressed. Now, consider the Millian propositions associated with the sentences \( \forall x (F(x) \rightarrow H) \) and \( \exists x (G(x) \wedge H) \). I might believe that the F is H even if I do not believe that the G is H and even if, unbeknownst to me,
the F is the G. In this case I might believe contradictory M-propositions but I would not be epistemically censurable on anyone’s theory. As I have said, Millians deny that sentences containing definite descriptions express Russellian proposition. But they cannot deny that being belief-related to one and a contrary M-proposition is not in itself irrational. I am just claiming that the same is true when the M-propositions are expressed by sentences containing proper names. There is simply no acceptable notion of M-irrationality.

Millianism should instead take a sort of mixed position: admit that the epistemic notion of F-rationality is the only kind of rationality but insist on the semantic notions of M-meaning, M-belief, etc. This mixed position also has available a subtle response to the argument against Millianism I have given above in Part I. Recall that, as I have understood them, the puzzling cases are situations in which ordinary rational people are, because of certain semantic principles, characterized as irrational. But the puzzles all involve an intermediate step characterizing these people as having contradictory beliefs. Now, a Millian who accepts only F-rationality could simply reject the definition of “rational” used in the penultimate step of the traditional argument and of the two analogs (see Part I above). That is, she could simply reject the claim that if S has contradictory beliefs, then S is not rational. Of course, that is a disadvantage of this position. After all, as I have argued, if it is not irrational occurrently and reflectively to hold obviously contradictory beliefs, what is? But remember that the puzzle cases only make it plausible to accuse the subject of having M-contradictory M-beliefs.

40 See above pp. 33–5.
41 This is the position that says I force a too-intimate connection between semantic and epistemic concepts.
The Fregean and this ‘mixed’ Millian could agree that having M-contradictory beliefs does not make you irrational: it does not make you \textit{F-irrational} and that’s the only kind of rationality. Thus the argument I gave earlier, insofar as it is urged as a \textit{refutation} (though in fact it is not so urged), is directed against a Millianism that holds that if you have M-contradictory beliefs (which are the only kinds of belief there are) then you are irrational in the only sense of the word. \textit{That} position does not give us the resources to say something we intuitively want to say: that whether you are rational is a function of the contents of your beliefs and that you are responsible for avoiding \textit{contradiction} in a way that is significantly different from the way that you are responsible for avoiding \textit{error}. A lesson of the traditional argument and Kripke’s analogs, as I take them, is precisely that \textit{the notion of contradiction delivered by classical extensional logic is not apt for use in the characterization of the epistemic notion of rationality.}

Associated conceptions of logic

Associated with the two theories I have discussed are thus two general conceptions of logic. Remember that, because they have different understandings of prior concepts (such as \textit{ambiguity}), the two theories have different understandings of notions such as \textit{contradiction, validity, implication}, and so on. In effect, Millianism corresponds to classical extensional logic while Fregeanism will correspond to some intensional (or perhaps hyper-intensional) logic.

The correspondence I have in mind here derives ultimately from the fact that in extensional logic as in Millianism, the only semantic value (in the sense of what is assigned to it by an \textit{interpretation}) of a logical constant (the formal analog of the name) is a referent. The truth value of the well-formed formulae of an extensional logic are taken to be determined by \textit{inter alia} the assignment of referents to the logical constants.
Similarly, the truth values of the grammatical sentences, according to Millianism, even though these sentences may contain intensional operators like ‘believes,’ will be determined by *inter alia* the assignment of (no more semantic value than) referents to the names. With respect to names, on the Millian theory, these operators do not deserve to be called “intensional” because they operate only on the *extensions* of the proper names in their scope.

By contrast, in Fregean theories, a proper name is assigned, basically, a sense, a kind of intension, which determines its extension. Intensional operators now operate on those meanings. Similarly, in intensional logic, the assignment of extensions to the sub-sentential symbols is not sufficient to determine the truth value of the expression as a whole. What makes the situation confusing, however, is if we say, as perhaps we should not, that in intensional logic the *reference* of a term is its *intension*.

I would suggest then that the difference between the two semantic theories under consideration can be connected with two general conceptions of the nature of logic (which may itself underlie the distinction between intensional logics and extensional logics). One way of capturing the different general conceptions is as follows:\(^{42}\) logic as *the physics of arbitrary objects* (from Gonseth’s “la physique de l’objet quelconque”) and logic as *the model of the state of mind of the arbitrary rational thinker*. Another way to put the point (and to put it controversially): there is something right in the hoary view of logic as “rules for the direction of mind.”

The characterization I have given so far, of the two theories’ underlying perspectives on the nature of logic and semantics, is simplistic and must be qualified. I mean the characterization primarily to be suggestive. To avoid misleading, however, I

\(^{42}\) In an attractive formulation due to Bas van Fraassen.
will say that it is not the case that Millianism makes no room for the role of language in representing our thought. Millianism has a clear notion of the content of belief: M-content. But that notion of content derives from Millianism’s perspective on the nature of semantics in the following way: the M-content of a belief is exhausted by its representing a certain part of the world as being a certain way. This is a useful notion of content to have. As I have said it does allow us to make important judgments concerning an agent’s mental state (but then, even t-belief allows to do that).

Thus, Millianism does characterize a notion of content that can be understood as representing our thought. Recall that an agent who has M-contradictory M-beliefs is in an epistemically imperfect state. He is not properly subject to logical censure (his M-contradiction may be the result of unavoidable ignorance—like Pierre’s) perhaps, as would someone who had F-contradictory F-beliefs, but M-contradictory M-belief guarantees error. Insofar as the M-contents of a subject’s M-beliefs determine how his mind represents a particular part of the world as (partly) being, that kind of belief is significant.43

Nevertheless, the guiding idea of Millianism still seems to me to be that language is a vehicle for representing certain parts of the world as being certain ways. That semantics will yield a content for any belief, but only inasmuch as it yields a content for any sentence embedded in a belief attribution. And the content it yields for those embedded sentences is a matter of, so to speak, configurations of extensions.

The discussion so far has emphasized the intimate connection between M-contents and what is said by the utterance of a declarative sentence and the intimate connection between F-contents and what we have in mind in holding a propositional

43 Cf. the notion of t-contradictory t-belief.
attitude. This might suggest two further claims that I will close by denying. First, on my view it is not the case that utterances of declarative sentences do not also express, as a semantic value, a Fregean proposition. Furthermore, there is no reason to deny that the contents of some thoughts might well be exhausted by nothing finer than a Millian proposition. Indeed, if, as seems phenomenologically plausible, some of our thoughts consist of nothing more than some ordinary language ‘rushing before the mind’s eye’ (a serviceable if fatigued metaphor), then these thoughts may have no more than Millian content (though they might also be understood as having a metalinguistic, Fregean, content\textsuperscript{44}).

My position is that an ordinary assertoric utterance of a declarative sentence (containing a proper name) should be understood as having two kinds of meaning: a Millian content which consists in the state of the world that must exist (contrast ‘obtain’) for that sentence to be true, and a Fregean content, which consists in the thought expressed by the subject that uttered that sentence. And in some cases, the content of a cognitive attitude may be just a Millian content (especially in cases where one is thinking by talking to oneself in a natural language).

A second claim that might be suggested by the discussion so far, but which I would deny, is that the Millian content of an utterance of a declarative sentence is determined by the Fregean content of that utterance. I see no good reason to make this further claim. Making it would take us closer to Frege’s position; but we would inherit some of the well-known difficulties of that position. Indeed, the denial of this claim is a central difference between the overall position I am defending here and Frege’s.

\textsuperscript{44} For a development of such a position, see Katz.
A familiar example will help to indicate my resistance to this additional claim: In *Naming and Necessity* (pp. 83–4) Kripke imagines that Gödel did not discover the incompleteness proof but stole it from a little-known mathematician named ‘Schmidt.’ Kripke imagines further that an agent, who has nothing more in mind in using the name ‘Gödel’ than *the discoverer of the incompleteness of arithmetic*, says something like, “Gödel was brilliant.” Now, according to the analysis I have given so far, the agent’s utterance of “Gödel was brilliant” exhibits two kinds of semantic value. First there is the Millian content of the utterance. That consists (we can suppose) of an ordered pair of Gödel and brilliance.

Second, there is the Fregean content of the utterance. That consists of an ordered pair of the sense of “Gödel” and *brilliance*. But if we have supposed that the sense of “Gödel” is just *the discoverer of the incompleteness of arithmetic*, then the Millian proposition that is determined by the Fregean proposition expressed by the agent’s sentence will consist of an ordered pair of *Schmidt* and brilliance. It seems clear, however, that the agent refers *linguistically* to Gödel, and does not refer linguistically to Schmidt, in uttering “Gödel was brilliant.”

The agent *mentally* represents Schmidt in uttering “Gödel was brilliant” and his utterance carries the propositional content that underwrites that mental representation (that is, the utterance expresses the Fregean proposition consisting of the ordered pair of *being the discoverer of the incompleteness of arithmetic* and *brilliance*). But the other kind of propositional content carried by the utterance, the Millian proposition consisting of Gödel and brilliance, is not determined by the Fregean propositional content of the utterance. No; the linguistic content of a given declarative utterance is, we can suppose, determined by a complex causal condition stretching back in time and out in space to other users of the language.
Chapter 2: Irrationality in Externalism

The issue of Millianism and Fregeanism in philosophy of language and the issue of externalism and internalism in philosophy of mind are deeply related. It will emerge that the relation is more complex than might be thought. Understanding their relation clearly will enable a useful philosophical insight: in this chapter, I will present a criticism of externalism in philosophy of mind which is an extension of an objection I have made in Chapter 1 against Millianism in philosophy of language.

A central claim of Chapter 1 was that agents who represent the same parts of the world as being the same way, *i.e.* agents who have the same Millian (or neo-Russellian) contents of thought might still be different in a significant way: one might be rational while the other is not. A theory of propositional attitudes that limited itself to Millian contents would ill-serve the purpose of understanding an agent’s rationality in terms of relations among the contents of her thoughts. I claimed that we need to supplement the Millian theory with some kind of Fregean theory to capture this intuition.

Here I will claim, similarly, that by making the contents of an agent’s thoughts a matter that is not determined by his overall internal state, externalism faces an analogous problem. Agents who are internal duplicates can differ in a significant way: one might be rational while the other is not. But just as agents with the same thought contents should not differ with respect to rationality, agents who are internal duplicates should not so differ. By making it possible for internal duplicates to have different thought contents (because their thoughts arise in—or as a result of—different environments), externalism will violate the intuition that logical coherence (and, to that extent, rationality) is basically a matter of *internal* consistency.
This result will emerge toward the end of this chapter as a corollary to another result. First I will show that externalism cannot maintain the natural, and naturally sharp, distinction between ignorance and incoherence. By making the determination of mental states a matter partly external to the agent, externalism opens up the possibility that mere empirical ignorance will be sufficient to make an agent logically incoherent. This will be shown by means of an example that I will develop in detail below. It will then be possible to see how the phenomenon in question is independent of the agent’s overall internal state: the logically incoherent agent(s), in the example, could be internally just like other agents who (because they do not have the relevant empirical ignorance) are not logically incoherent.

Before proceeding to the example and the corollary, however, it is worth making externalism itself more explicit and getting clearer about the relation between the issue of Millianism and Fregeanism in philosophy of language and the issue of externalism and internalism in philosophy of mind.

Part I: Two Controversies

What is externalism?

Millianism is a position about the semantic value of proper names. It is a position in the philosophy of language. How, exactly, is it related to issues in the philosophy of mind?

If we accept the thesis of relationality (that belief ascriptions report relations between believers and propositions), then Millianism has implications for the truth value of ordinary belief ascriptions. Whether S believes that P, for example, will be equivalent to whether the sentence that ascribes him that belief (the belief that P)—i.e., whether “S believes that P”—is true. And the truth value of sentences, including belief ascriptions, is
partly a matter of their semantics. Therefore, in a way, Millianism has implications concerning who believes what. This is a surprising result.

S believes that p if and only if “S believes that p” is true. But whether “S believes that p” is true depends on the semantics of that sentence. And, significantly, when the substitution instance for ‘p’ contains a proper name, Millianism and Fregeanism may provide different truth values for the embedding sentence (the ascription as a whole). Where Fregeanism has the result that “S believes that p” is false, Millianism may have the result that it is true. So whether S believes that p is related to the semantics of proper names.

I have argued that a division of the notion of proposition makes this result less mysterious. In making an ascription, we use an expression that has two kinds of semantic value—that expresses two kinds of proposition. Millianism may be right about one of those kinds of semantic value: linguistic representation. Just focusing on that kind of semantic value for a moment, if an agent believes that a is F, and if a=b, then the agent believes that b is F. This is an implication of Millianism. But this sense in which the agent believes that b is F is equivalent to that agent’s believing, of b, that it is F. It may be that the kind of belief about which Millianism has implications is perhaps a less worthy deserver of the name ‘belief’ than the state that goes by the same name (‘belief’) in the Fregean theory.

Now, I have claimed that Millianism in the philosophy of language and externalism in the philosophy of mind are intimately related. But more than one position in the philosophy of mind goes by the name of ‘externalism’; so it is worth making explicit exactly what I will mean by that term.

To begin with externalism makes a stronger claim than the weak thesis that some intentional properties are relational properties—properties defined in terms of relations
between an agent and what that agent thinks about, where what the agent thinks about may be something external to that agent. That weak thesis might be accepted by any theorist, internalist or externalist. In having a belief about something, one is thereby in some kind of relation to that external thing. It is a confusion to understand the internalist as denying that we can be related, by belief, to external objects.

What then does externalism claim? It turns out to be surprisingly difficult to characterize it precisely. Putnam announced, famously, that “meanings ain’t in the head.” But what is the literal content of the metaphor? It appears to be the denial of an internalist claim that meanings are in the head. We will understand externalism better by understanding how internalism can be seen as the claim that meanings are in the head.

Internalists believe that an agent’s mental states are in some important sense hers. Unlike for example the property of being a brother, which essentially involves a relation to an external entity, believing that you are a brother, according to internalism, is an internal property of any agent that has that belief: it may be a relational property, but any relations involved must be wholly internal to the subject. Any two individuals in the same internal state must have the same mental states. The externalist, by contrast, holds that some mental states are constituted by the subject’s standing in a relation to an external object. Thus, according to externalism, internal duplicates who are differently situated may have different mental states.

So far we have an internalist claim about mental states. How does this become a claim about meanings? Among mental states there is an interesting subclass of states distinguished by having a content. So, for example, it is impossible to have a belief without believing that something is the case—without believing that $p$. The content of the belief is then simply that $p$. This feature characterizes the class of mental states that are called ‘propositional attitudes.’
Contrast anxiety. Anxiety is a mental state; but it is possible simply to be anxious—without being anxious that something is or will be the case (though of course it is also possible to be anxious that something is or will be the case). Anxiety, unlike belief, does not necessarily have a content and thus is not generally a propositional attitude (though some instances of the type may have propositional content).

Consider now the propositional attitudes. When we ascribe one of these attitudes to an agent, *e.g.* by saying, “Jones believes that Cicero denounced Catiline,” we use a sentence that embeds another sentence. That embedded sentence—in this case, “Cicero denounced Catiline”—is an essential ingredient in the ascription: it seems to express precisely what the agent believes. This encourages a traditional understanding of the nature of such ascriptions: a belief ascription reports a relation between a thinker and the content of that thinker’s belief. The thinker is referred to by the grammatical subject of the ascription and the content is expressed by the embedded sentence. Finally, on this traditional understanding, what a declarative sentence expresses is precisely its meaning.

This traditional view of the nature of belief and of belief ascription helps explain how it is that the internalist claim that the contents of mental states are determined internalistically can be opposed by the externalist slogan “meanings ain’t in the head.” The internalist believes that the contents of mental states are determined independently of any entities external to the subject of those mental states. However, if propositional attitudes consist of an attitude toward what is expressed by a declarative sentence (as on the traditional understanding mentioned above), then the contents of these mental states, the propositional attitudes, are determined with respect to the meanings of sentences (in particular, the meanings of the sentences used to report those attitudes).

Thus the internalist must hold that the meaning of a sentence embedded in a true belief ascription must in some sense be ‘in the head’ of the subject of the belief ascribed.
To say that meanings ain’t in the head would be to oppose this internalist position. If meanings weren’t in the head, then the contents of some mental states, the propositional attitudes, would be determined only with respect to entities that are external to the subject of those attitudes, namely, the propositions that constitute the contents of those attitudes. It is important to recognize clearly this relation between the externalist slogan about meanings and the philosophical position about mental states that it opposes.\footnote{The internalist needs to make an exception for what we might call \textit{indexical} mental states: each of two internal duplicates might believe that \textit{he} is in London. But it may be that only one of them is. If \textit{what} they believe is one and the same thing, then what they believe is both true and false. So they do not believe the same thing. Characterizing indexical mental states precisely is a difficult matter that I cannot undertake here; but the exception should be kept in mind in what follows.}

So externalism amounts to the claim that internal duplicates can have different propositional attitudes. Putnam’s early paper on this subject has an example that begins to make this position plausible. He imagines a possible world very much like the actual world except that in this possible world there are two earths. The only difference between planet Earth and its twin (in this possible world) is that on ‘Twin-Earth’ the stuff that flows in rivers and streams and falls from the sky as rain, \textit{etc.}, is not H2O but another substance with the chemical composition XYZ (which we can call ‘T’water’). We are to suppose further that on Twin-Earth there is a perfect internal duplicate of someone on Earth (ignore the internal difference that results from the presence of water in our bodies).

The externalist invites us to the following intuitions. When the agent on Earth has the thought that he would normally express as “water is wet,” the twin on Twin-Earth would have a \textit{different} thought, though one he would also express as “water is wet.” According to the externalist, it is true of the agent on Earth that he believes that water is wet; but it would be false of the agent on Twin-Earth that he believes that water is wet.
The agent on Twin-Earth has never come across any water in his life; how can he have any beliefs about water? The agent on Twin-Earth, the example continues, has beliefs about T’water—the substance XYZ. The two agents, though internal duplicates, differ with respect to propositional attitudes: one believes, and the other does not, that water is wet. They would both say the words, “water is wet”; but when the twin on Twin-Earth says those words he is referring to T’water—the stuff that he drinks and swims in—and not to water, with which he has never had any contact. In saying the words “water is wet” the twin would not be expressing the belief that water is wet, but the belief that T’water is wet.

This is Putnam’s early Twin-Earth example and the intuitions about it we are invited to share. If we share them all, there is no choice but to bid farewell to internalism: it is incompatible with internalism that there be a possible world in which two agents are internally alike but differ in their mental states. And it must be admitted that the example is compelling. Whatever our considered opinion of the relevant positions, the intuitions we are invited to share are attractive and seem natural.

Externalism, then, is a view about the determination of the contents of certain mental states. The mental states in question are perhaps best specified by a list. Any attempt to categorize them may confuse issues that will arise in later theorizing. So, for example, although one way to characterize the relevant mental states is by categorizing them as propositional attitudes, where propositional attitudes are understood as constituted by an attitude and a proposition, this criterion would prejudge a question that some would controvert: that propositional attitudes are not to be individuated by the propositions that are their contents. It is perhaps better to rely on an intuitive notion of
What it is to be the content of a mental state is a controversial issue. It is defined in different ways by different theories. These definitions do, however, agree on the following characteristic: the content of a belief, desire, hope, or fear is what you believe, desire, hope or fear. If what I believe is that it is raining, then that it is raining is the content of my belief. If my hope is that you should come to dinner, then that you should come to dinner is the content of my hope; and so on.

Now we are in a position to approach externalism’s constitutive thesis:

(Externalism) The contents of an agent’s mental states are not determined except with respect to relations between that agent and entities that are external to that agent.

Externalism claims that the contents of a given subject’s mental states are not determined even given a full determination of his internal (physical) state. Two subjects can be in the same internal state and still differ in which mental states they are in. The internal physical nature of a subject does not determine his mental nature. It is possible that two individuals in the same internal state have different mental states. To put externalism in the form of a slogan—the mental does not supervene on the internal.

Does Millianism entail externalism?

Of course, even a list might seem to prejudge certain questions since by including some states and excluding others we make an implicit and perhaps controversial choice as to which kinds of states should be grouped together.
I have claimed that externalism is intimately related to Millianism. But it could be made still clearer how they are intimately related. After all, Millianism is a thesis about the semantic value of proper names—that they have no semantic value other than their bearer. Externalism, on the other hand, is a thesis about the individuation of mental states.

Now recall the following popular assumptions: (i) the semantic value of a sentence is a proposition, (ii) mental states (the kind of mental states on our list) consist of an attitude toward a proposition. (ii) amounts to the thesis that, within a given attitude type (say, belief), mental states are individuated by their contents. If contents are propositions, and if propositions are constituted partly by objects that are external to the subject of the mental state (since proper names can refer to objects external to the subject), then mental states can be individuated with respect to entities external to the subject of those states. To believe that Cicero denounced Catiline, for example, according to Millianism, is to stand in the belief relation to a proposition constituted in part by Cicero himself. It is easy, but not for that ease any less incorrect, to slide from this point to viewing Millianism as an instance of Externalism. It is not.

First, just because we can individuate an object with respect to entities that are external to that object does not mean that we cannot also individuate it internalistically. My brother is my parents’ youngest child. Whether anyone is identical with the person that is my brother is a matter that is determined by whether that person is the youngest child of my parents. So it is a matter that is determined with respect to facts about my parents. However, there are ways of individuating my brother (by appeal to his genetic code perhaps) that do not appeal to entities external to him. So whether anyone is the person that is my brother is a matter that is determined independently of facts about my parents.
Second and more importantly, if whether you stand in the belief relation to some Russellian proposition is determined by whether you stand in some other relation to something other than a Russellian proposition, e.g. a proposition-like entity composed of items entirely internal to you, then internalism would be true; but so too could Millianism.

This is a difficult point to make so I’ll spend some time clarifying and emphasizing it. To begin with, recall that Millianism is a thesis about the constitution of the contents of cognitive attitudes. Externalism, on the other hand, is a thesis about the individuation of cognitive attitudes. If we take it for granted that (within types, such as belief) cognitive attitudes are individuated by their contents, i.e. that two belief tokens count as of the same belief type if and only if they have contents of the same type, then it is easy to slip into the view that Millianism entails externalism. It does not.

Consider the following argument: (i) the contents of our cognitive attitudes are constituted by objects that are external to us; (ii) our mental states are individuated with respect to their contents; therefore, (iii) our mental states are individuated externalistically, with respect to the entities that constitute the propositions that individuate our mental states. This argument is sound as far as it goes, but (iii) is insufficient for externalism. Externalism insists that mental states cannot (in general) be individuated except with respect to external objects.3

---

3 It might be thought that given the notion of individuation that I have defined, externalism is obviously wrong. Surely there have been some beliefs that have only ever been had by a single person. If all that is required for individuation is the material determination of the mental state by the internal state, then any one of those beliefs will be individuated by the internal state of the agent that had it. Of course; but externalism doesn’t deny that there are some mental states that are individuated by some internal states—what externalism denies is that all mental states can be individuated by the internal state of the agent that has them.
For this kind of argument to yield externalism as a consequence, then, (ii) would have to be amended to read as follows: (ii') our mental states cannot be individuated except with respect to their contents. But this is a stronger thesis than any Millian need accept. Millians would and certainly can accept this strong principle; but they need not.

To see what is at issue, compare (ii') with the thesis of relationality mentioned earlier: belief ascriptions report relations between believers and the propositions expressed by their (the ascriptions’) complements. The thesis of relationality encourages the picture that the only relevant dimension of variability within a given attitude type is content. A consequence of relationality is that if (a) S believes that Q and (b) the proposition that P = the proposition that Q, then (c) S believes that P. It seems then that what determines whether a subject S believes that P is precisely whether the following substitution schema is satisfied: the proposition that P = the proposition that Q and S believes that Q. Indeed, the satisfaction of that schema is both necessary and sufficient for S’s believing that P. But that S believes that P may also be determined by factors that are entirely internal to S.

Suppose for example that the Millian viewed the belief relation as the existential generalization of another relation. This other relation, call it ‘B,’ could be a relation between believers and proposition-like entities (use the variable ‘i’ to range over them) constituted entirely of items internal to the believer (they could be something like senses). Now these proposition-like entities might have systematic relations to propositions—even when propositions are understood on the Millian model. So, for example, there might be a function D from the i’s to propositions. Accordingly, the Millian might hold that whether S believes that P is determined by whether there is an i such that B(S, i) & D(i, P). Whether there is an i such that B(S, i) is a matter that is entirely internal to S. And the (so to speak) Millian proposition believed by S is determined by the i to which he stands in the B-relation. On this view then, fix the
internal entity to which an agent stands in the B-relation and you have determined, relative to a possible world, that agent’s belief state.

In this way the Millian could hold that belief is a relation between believers and propositions, that propositions are constituted *inter alia* by objects external to the agents who stand in the belief relation to those propositions, *and* that it is possible to individuate mental states with respect only to states and entities that are internal to the agent of those states. In asserting that belief ascriptions report relations between believers and propositions we imply that mental states can be individuated by their contents. But relations can be determined by other relations. And if the relation asserted in a belief ascription is determined by another relation, one that is itself determined by states and entities entirely internal to the believer, then mental states can also be individuated by this deeper relation. *Millianism does not entail externalism.*

I highlight this point in order to bring out certain aspects of the position defended above in Chapter 1. I have claimed that Millianism and Fregeanism are not incompatible theories of the same phenomena but rather independent theories of different phenomena. I have combined this claim with a general distinction between mental representation and linguistic representation. The need for the distinction over and above the claim of independence can now be made clearer.

Consider Frege’s original position. He distinguished two kinds of semantic value for names. Millians hold that names have only one kind of semantic value. Of course these positions are at odds only if the term ‘semantic value’ is used univocally in the two positions. In order to get some purchase on the notion of *semantic value* we can stipulate that the semantic value of a declarative sentence is the proposition expressed by that sentence. Now, however, the very phenomena that led a Fregean to distinguish two kinds of semantic value for names might lead him to distinguish two kinds of thing to be the
In response to the Millian, a Fregean might allow that the so-called ‘proposition’ expressed by a declarative sentence (containing a name) is a structured entity composed partly by the referent of the contained name; and he might even allow the identification of the semantic value of a declarative sentence with the proposition expressed. Thus he would allow, in a sense, that the only semantic value of a name is its reference. But he would insist that there is another feature of names that determines that semantic value. And, correspondingly, there is another feature of sentences that determines the proposition expressed. This other feature of sentences is also a structured entity not unlike the proposition expressed.

Of course, sense determines reference only relative to a possible world. So this putative Fregean Millianism might seem to be an externalist view after all. In order to determine the agent’s mental state, we need to fix not only the Fregean proposition, composed entirely of senses, to which the agent is related, but also the world around the agent. Thus, it might seem, this kind of Fregeanism would not be compatible with an internalist Millianism at all. But this response misunderstands externalism’s constitutive claim.

The contrast can be brought out more clearly in the following way. The internalist thesis that internal state determines the contents of mental states is meant to have modal force. That is, the internalist holds that necessarily if some agent has a mental state with content c and has overall internal physical state s, then any agent with overall internal physical state s has a mental state with content c. But although it has modal force, the thesis is still, in a certain sense, an “intra-world” claim. What it claims is only that for any possible world w, if some agent in w has a mental state with content c (in w) and has

\textit{proposition} expressed by a declarative sentence (restricting our attention to sentences containing proper names).
internal physical state s (in w), then any agent in w with internal physical state s (in w) has a mental state with content c (in w). This can, but should not, be confused with what we might call a “cross-world” version of the thesis. On the cross-world version, the internalist claims that for any worlds w₁ and w₂, if some agent in w₁ has a mental state with content c (in w₁) and has overall internal physical state s (in w₁), then any agent in w₂ with overall internal physical state s (in w₂) has a mental state with content c (in w₂).

It is instructive to compare this distinction to a distinction among varieties of supervenience. Property type A “globally” supervenes on property type B just in case two worlds identical with respect to the distribution of B-properties must be identical with respect to the distribution of A-properties.⁴ So, for example, the general claim that the mental supervenes on the physical is true just in case any two worlds that are physical duplicates must be duplicates with respect to mental properties.

Using that understanding of supervenience, we could say that the contents of an agent’s mental states supervene on the overall internal physical state of the agent if and only if, necessarily, any two agents that are duplicates with respect to overall internal physical state must be duplicates with respect to contents of mental states. But this is just the strong, “cross-world” internalist thesis that does not properly characterize the dispute between contemporary internalism and externalism. Putnam’s and Burge’s examples are striking because if they were successful they would refute even the weaker “intra-world” internalist thesis. The internalist Fregean needn’t affirm the strong cross-world version of the internalist thesis. He can allow that an internal duplicate of yours might have (in another possible world) different thought contents.

⁴ Global supervenience is equivalent to “strong” supervenience. See Kim (1984, p. 168). Later, however, Kim (1987) holds that global supervenience entails neither strong nor weak supervenience and is too weak a concept for a proper formulation of materialism. We cannot consider this controversial topic here.
What I have called the “intra-world” internalist thesis, on the other hand, corresponds to so-called weak supervenience. Property type A weakly supervenes on property type B if and only if necessarily, duplicates with respect to B-properties are duplicates with respect to A-properties. In his discussion, Kim points out that weak supervenience “requires only that within any possible world there not be two things agreeing in B but diverging in A…. It does not require that if in another world an object has the same B-properties that it has in this world, it must also have the same A-properties it has in this one.”

Many physicalists would hold that the mental strongly supervenes on the physical. Fixing the distribution of the physical properties in a world will fix not only the internal but also the external relational physical states of the agents in that world. Two worlds identical with respect to all physical properties must be identical with respect to the distribution of the mental properties in those worlds. Weak supervenience is too weak to represent the physicalist position since it would allow worlds that are physical duplicates to differ with respect to their distribution of mental properties.

But that issue is simply independent of the internalism/externalism dispute. When it comes to the supervenience of the contents of our mental states on our internal physical state, we need hold nothing stronger than weak supervenience. Some internalists might have held out for strong supervenience even here—even if the modality is nomic rather than metaphysical. That is, some internalists might have held that it is nomically necessary (even if not metaphysically necessary) for duplicates with respect to overall internal physical state to be duplicates with respect to the contents of their mental states. But Putnam’s and Burge’s examples cause difficulty even for the weak supervenience thesis. Those examples concern individuals who are overall internal physical duplicates.

---

within a possible world. If the examples were successful, even the claim of weak supervenience would have to be abandoned.

Thus the fact that sense determines reference only relative to a possible world does not undermine the point that Millianism is compatible with internalism—properly understood. The Millian could agree with the internalist that an agent’s mental state is determined, for any possible world, by her internal state.

The issue is whether in specifying an agent’s internal state we need to specify, so to speak, which agent is in that internal state. Internalism holds that such a further specification is irrelevant: whoever is in that internal state will have the given mental state. Externalism, on the other hand, claims that the internal state by itself will not determine which mental state a given agent is in. That would require a specification of that agent’s relational properties. These properties will be determined by the general physical description of the given possible world. But that general description will not necessarily determine a unique set of relational properties for a given internal state. Therefore, according to the externalist, a general physical description of the possible world, together with a specification of an internal state, is insufficient to determine the mental state of any agent in that internal state: the relevant agent’s relational properties must also be specified. And different agents in the same internal state may have different relational properties.

Two views of the constitution and determination of propositional content

Now, the kind of internalist Millianism I have been describing would really transform the issue between Millians and Fregeans into a spurious dispute. And in fact the dispute between Millians and Fregeans is not best understood in these terms. Contemporary Millians do not accept that there is some other proposition-like entity, composed entirely
of items internal to the believing agent, that determines—relative to a possible world—the proposition believed. Instead of the B-relation, as I have characterized it above, modern Millians hold that belief is the existential generalization of something they call the BEL-relation. This BEL-relation is a three place relation holding between believers, Russelian propositions, and ways of taking those Russelian propositions.

The internalist Millianism discussed above was a piece of theoretical construction—to make explicit the point about constitution and determination. One way to see the issue between Millians and Fregeans is to consider whether the third term of the BEL-relation determines the second. If it did, the relation would be redundant and should be replaced by something like the B-relation. But if it does not—and modern Millians hold that it does not—then the issue between Millians and Fregeans does not degenerate into the terminological dispute sketched above.

To clarify: Millianism and Fregeanism are at odds over the constitution of propositions expressed by declarative sentences containing proper names. The issue of internalism and externalism is a dispute over the determination of the propositional contents of certain mental states. I have shown that Millianism is consistent with internalism. However, an internalist Millianism would not be fundamentally at odds with Fregeanism—the dispute would be purely terminological. Linguistic legislation could bring about a satisfactory accord.

Interestingly, I argued in Chapter 1 that Millianism and Fregeanism are not deeply at odds. And I have argued that terminological legislation could bring about a satisfactory accord. But I deny that the dispute between Millianism and Fregeanism is purely terminological. The tension between them is due to their attention to different domains of discourse each of which (properly) constrains the notion of content. This leads to the equivocation that I claim underlies their apparent opposition. An internalist
Millianism of the kind discussed here, on the other hand, need not differ from Fregeanism even with respect to explanatory focus. In other words: though on my view the compromise between Millianism and Fregeanism is achieved at the cost of distinguishing what the theories are primarily about, an internalist Millianism would be compatible with Fregeanism even if we took their subject matters to be the same.

Again, an internalist Millianism, though a theoretical possibility, is a mere terminological variant of Fregeanism (save for a technical detail to be discussed below). However, the issue between Millianism and Fregeanism is not purely terminological. Their apparent dispute is the result of wholesale equivocation. But that wholesale equivocation derives from the discrepant demands placed on a set of theoretical terms by two domains of phenomena. That is, the phenomena that Millianism and Fregeanism are attempts to explain fall naturally, I have claimed, into two domains. These domains place competing demands on any theory that attempts to explain the relevant phenomena. There are simply two kinds of thing that it is to be a content or a proposition and hence two kinds of thing that it is to be a belief. The issue between Millianism and Fregeanism, according to me, is only derivatively terminological; its source is in their respective primary explanatory purposes.

But now, if I am right that an internalist Millianism would be entirely compatible with Fregeanism, even for a given range of explanatory purposes, then the dispute between Millianism and Fregeanism, across the range of explanatory purposes (in other words, abstracting away from the explanation of their opposition that I have given in terms of explanatory focus) reduces to the issue of externalism. Setting aside positions that would make Millianism and Fregeanism compatible amounts to taking seriously the controversy between internalism and externalism and insisting that Millianism be externalist in a way that Fregeanism is not.
What is interesting for our purposes, then, is that what deep dispute there is between Millianism and Fregeanism is, with a qualification to be discussed presently, simply the dispute between externalism and internalism. It is only by disallowing the terminology that would make internalism compatible with Millianism that we maintain Millianism’s opposition to Fregeanism.

I will begin to explain this point by making the technical clarification that I have promised. I have claimed that an internalist Millianism would be compatible with Fregeanism. This requires a qualification. Frege claimed that senses, the grasping of which constitutes thought, are objective, mind-independent entities. They are not internal objects in the way of sense-data, for example, or of what Frege called “ideas.” They are in this way unlike images, at least on some understandings of imagery. So for the Fregean too, thoughts are at least constituted by external objects. By now we know better than to assume that this implies externalism. So the question arises, “according to Fregeanism, does the internal state of an agent determine which senses he is grasping?”.

This is a question that has not, it seems to me, received the attention it deserves. It is easy to assume that Frege’s view was that images stand to senses in the way that senses stand to references. That is, it is easy to assume that image determines sense. This would transform Frege into an internalist. But I am skeptical that Frege is best understood in this way. In fact, it may be that Frege took the grasping relation to be a basic relation—a relation that does not in general supervene on any physical state of affairs (at all, much less the physical state internal to the agent).

So it is at least compatible with what Frege did say that he took a specification of an agent’s internal state to be insufficient to determine the senses to which that agent stood in the grasping relation. In order to determine the agent’s thought, on this view of
Frege, it is necessary to specify the agent’s relations to external objects; in particular, it is necessary to specify which senses the agent is grasping.

I will assume that Frege was in this sense an externalist. But it is important to distinguish Fregean externalism from the brand of externalism that will be the main target of the rest of this chapter. Although Frege (plausibly) held that mental states cannot be individuated except in terms of relations to external entities, those external entities relations to which partly determine an agent’s mental state are all accessible to that agent in a special way. Senses are *graspable*. We are acquainted with them. They are directly accessible. And that distinguishes Fregean externalism from the kind of externalism that I will discuss.

The externalism that I will discuss is the position by that name that has become increasingly popular in philosophy of mind. And this position is characterized by the feature that the entities relations to which are essential to an individuation of an agent’s mental state are not only external to that agent but are psychologically remote as well. It is facts about how the agent is embedded in his physical environment that, according to the externalism at issue here, are essential to determining the agent’s mental state. And these facts are not immediately available to the agent in the way that the facts about which senses he is grasping would be.

To eliminate the need for constant qualification, in order to separate Fregean externalism from the externalism under discussion here, it is perhaps worth amending the earlier definition of externalism so as to exclude Fregean externalism explicitly. The following definition would make Frege an internalist (so long as internalism is simply the negation of externalism, as I have assumed throughout):

\[(\text{Externalism}') \quad \text{Mental states must be individuated with respect to relations between any subject of those mental states and entities that are external and not directly accessible to that subject.}\]
Though controversial, externalism has grown in popularity over the last two decades so that it can now be taken almost as orthodoxy. I believe, however, that the position is deeply flawed. In what remains of this chapter, I will present a difficulty for externalism.

Part II: A problem for externalism

According to me, externalism is incompatible with a deeply held intuition about epistemic responsibility. There is a basic difference between ignorance and incoherence. It is impossible that ignorance should be sufficient to make an agent incoherent. No matter what an agent’s state of empirical knowledge, that agent is in principle in a special position to maintain the logical coherence of her beliefs and other mental states. I will argue that externalism is not compatible with this deeply held intuition.

So my purpose here is not to demonstrate an escape from the externalist challenge. I propose to grant, for the sake of argument, the externalist examples together with the intuitions we are invited to share. Instead, I want here to demonstrate an unacceptable consequence of externalism. If we go ahead and adopt the externalist position concerning the individuation of mental states, then, I claim, we must give up a natural and intuitive view about rationality.

Two characteristics of the intuition stand out: first, it is an epistemological intuition, an intuition about kinds of epistemic censure. It is not, in itself, an intuition in philosophy of mind—in appealing to it, I do not obviously beg the central question. Second, the intuition is one even externalists should embrace. Indeed, no less an externalist than Saul Kripke has explicitly embraced an intuition very much like it: “…anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them. Precisely for this reason, we regard individuals who
contradict themselves as subject to greater censure than those who merely have false beliefs.”

Rationality, at least theoretical rationality, is a matter of having one’s beliefs in order. There are many ways in which an agent can fail to have her beliefs in order; there are many ways in which an agent can be irrational. This is a reflection of the fact that the order one’s beliefs must exhibit for one to be rational is multidimensional. That is, one’s beliefs must be in various different kinds of order. For example, justificatory relations among one’s beliefs will define one kind of required order. It is irrational to infer that the moon is made of green cheese on the basis of a belief that DNA is a double helix (assuming a normal set of background beliefs). In addition, rationality may necessarily involve certain kinds of epistemic procedures which could define another dimension of order among beliefs. So for example the epistemic failure of overgeneralization might be a matter of violating a required procedure of sufficient data collection; overgeneralization could be seen as a disorder between the belief constituting the overgeneralization and the paucity of beliefs that led to it.

There seems, however, to be one dimension of order that has a privileged position. An agent is somehow especially epistemically responsible for the simple logical relations among her beliefs. It is bad to form beliefs on the basis of insufficient evidence, or to base a belief on other beliefs that do not sustain it, but it is a distinctive epistemic failing to have logically incoherent beliefs.

I am talking here of occurrent, explicit beliefs. There are issues about one’s responsibility for dispositional or implicit beliefs. We are finite beings with limited attention spans—we cannot expect all of our beliefs to be closed under logical

---

implication, for example. It is not necessarily irrational to believe that \( p \), to believe that if \( p \) then \( q \), and to fail to believe that \( q \). But (i) if \( p' \) and \( q' \) are sufficiently simple for one to be able to entertain them explicitly, and (ii) if one believes \textit{occurrently} and \textit{explicitly} that \( p \) and that if \( p \) then \( q \), then (iii) if one explicitly entertains \( q' \), and does not change one’s mind about \( p' \) or about \( \text{if } p \text{ then } q' \), one must not believe that not \( q \). To take another simpler example, it would be irrational to believe that \( p \) and to believe that not \( p \) (again, assuming both beliefs are fully explicit and occurrent). In this relatively weak sense, we naturally give deductive inference a privileged position in our assessment of epistemic procedures—we expect any rational agent to be in principle in a position to avoid \textit{invalid} deductive inferences (by contrast to \textit{unsound} inferences which may indicate only ignorance). Thus, we take rational agents to be in principle in a position to recognize the deductive validity, by contrast with the soundness, of their reasoning.

Let me emphasize that I am making a claim about \textit{inference} and our responsibility for avoiding simple invalid inference (under certain constrained conditions). This is different from our responsibility for avoiding simple invalid \textit{arguments} (understood syntactically). Inference involves \textit{beliefs} and relations among them (in particular, among their propositional constituents). Syntactic arguments, as I understand them, are \textit{sentences} or sequences of sentences conforming to certain patterns. Our acceptance of sentences, however, is a matter that may be independent of our acceptance of certain beliefs. At least, we should not assume otherwise.

The fact that we all might sometimes accept as valid a sequence of sentences that is in fact an invalid syntactic argument does not show that we make invalid deductive inferences. It would have to be shown that in these cases the propositions we take those sentences to express themselves lack the requisite logical properties. Competence may not guarantee a perfect match between what we take the sentence to express and what it in fact expresses. I claim only that we have a special epistemic responsibility for the
simple logical relations among our beliefs. This distinction has the following result: If I accept a formal argument that contains an equivocation, there may nevertheless be no invalidity in the inference I am drawing. There may simply be a breakdown in the adequacy of the argument in representing my inference.

My claim will be that externalism is incompatible with this natural view of our special epistemic responsibility for the logical relations among our (explicit, occurrent) propositional attitudes. If mental states are individuated even partly with respect to entities that are external\(^7\) to the agent that has those states, then that agent cannot be found irrational (or even incoherent) simply because she cannot tell the difference between those of her inferences that are logically valid and those that are not. Indeed, on certain ways of developing externalism, an agent cannot be found particularly faulty if she explicitly holds contradictory beliefs or because she explicitly rejects a logical consequence of her other explicit beliefs. If externalism is correct, then the various kinds of logical incoherence are not especially blameworthy epistemic flaws; they are on a par with ignorance. But incoherence indicted rationality radically, ignorance at most mildly.

As a way of making my objection to externalism, I propose the following amendment to Putnam’s metaphor: “if meanings ain’t in the head, then neither is logical coherence.” But if logical coherence is not a matter internal to an epistemic subject, then he can be no more responsible for it than he is for other external epistemic relations—such as failing to see what’s before his eyes (a grave epistemic fault, no doubt, but different in kind from logical incoherence).

\(^7\) For present purposes, I ignore a subtlety concerning the possibility of direct acquaintance with external objects (such as properties). In effect I am not distinguishing, for the sake of argument, between external and not a potential object of direct acquaintance. But see above pp. 96–8.
Now, how is it that externalism will undermine the natural and intuitive view of rationality? I begin with an example. Consider the situation before the discovery of the difference between gold and so-called fool’s gold. These are substances which, though chemically different, are alike with respect to all ‘macro’ properties. Now, imagine a group of people who had extensive contact both with gold and also with fool’s gold. Because they could not distinguish the substances, they used the same term for both substances. In particular, they used the term ‘gold’ both for gold and for fool’s gold.

Suppose, however (and here I doubtless deviate somewhat from any historical facts), that all the gold they ever encountered came from deep within a particular mine. They believe that the substance is also found naturally on the earth’s surface but in fact gold is found only in the one mine. What they find on the earth’s surface is only fool’s gold. Now, I claim that according to externalism, not all of the thoughts they would express with the term ‘gold’ are about gold. Some of those thoughts, the thoughts that are in the right external relation to gold itself will be about gold; others, that are in the right external relation to fool’s gold, will be about fool’s gold. (Perhaps some other thoughts will have the right external relation to both and will thus be about both gold and fool’s gold but we will not be concerned with these thoughts in what follows).

To make this more plausible, suppose that some of the thoughts they would express with the term ‘gold’ are beliefs that are based on interactions had exclusively with the stuff they mine from the gold mine—interactions they had while mining gold. On the basis of these interactions, they come to have a belief that they would express as follows, ‘gold is G.’ It is reasonable to suppose that (according to externalism) this expresses their belief that gold is G. In saying ‘gold is G,’ these agents are related to gold in the way that the agent on Earth, in Putnam’s example, is related to water (or the agent on Twin-Earth is related to T’water). On the other hand, the agents are related to fool’s gold only in the way the agent on Earth is related to T’water (or the agent on Twin-Earth
is related to water). The agents’ claim is causally related in a significant way only to
gold. Any causal relations to fool’s gold are incidental to their claim that gold is G.

Now suppose further that these agents have some other thoughts which they
would also express with the term ‘gold’ that are based on interactions they have had
exclusively with fool’s gold—interactions they had while panning for gold in rivers. On
the basis of these interactions, they come to have a belief they would express as follows,
‘gold is F.’ It is reasonable to suppose that (according to externalism) this expresses their
belief that fool’s gold is F. This belief is based on interactions the subjects have had only
with fool’s gold—any causal relations to gold are incidental to this belief and to their
expression of it. Although they use the term ‘gold’ to express their belief, just as the twin
in Putnam’s example uses ‘water’ to express his belief, the belief is clearly (according to
externalism) not about gold but about fool’s gold, just as the twin’s belief is putatively
about T’water and not about water.

It might be thought that it’s unfair to externalism to attribute to it the
characterizations of these agents’ beliefs that I have given. Would externalism really be
forced to say, under the circumstances, (i) that the belief the agents express as “gold is G”
really is the belief that gold is G and (ii) that the belief the agents express as “gold is F”
really is the belief that fool’s gold is F? This challenge might be based on the idea that
the word ‘gold’ in the group’s language is like the word ‘jade’ in English. It might be
thought that their word ‘gold’ unambiguously refers to a motley collection of two very
different natural kinds, just as our word ‘jade’ refers unambiguously to the collection of
jadeite and nephrite. I have tried to design the example so that it is clear that their word is
ambiguous as between the kinds and does not refer unambiguously to the motley

8 Notice that treating ‘water’ in the Twin-Earth case in this way would undermine
that example’s support of externalism, as Putnam himself is well aware.
collection. But in any case, the externalist, as well as the internalist, will need some non
question-begging distinction between unambiguous reference to a motley class and
ambiguous reference to a pair of dissimilar natural kinds. Presumably, the satisfaction of
this condition could then simply be built in to the example.

Moreover and more generally, externalism must ultimately provide some kind of
external relation (likely causal), the satisfaction of which is necessary for reference. It is
unclear to me why the failure to satisfy that relation cannot also be appropriately built in
to the example. In other words, the central thesis of externalism is that the meaning of a
given word (on a particular occasion) is partly determined by its relation to the
environment of the user of that word. Call the relevant relation R. Now it seems to me in
general possible to create an example in which the agents’ use of ‘gold is G’ bears R to
gold and does not bear it to fool’s gold and in which the agents’ use of ‘gold is F’ bears R
to fool’s gold and does not bear it to gold. I contend that for many plausible such
relations R, the example I have given would do this already. But all that is necessary is
the general possibility. My example is only as unfair to externalism as is required by its
silence on the issue of which external relation is required for reference.

So the subjects in my example have two beliefs; and according to externalism,
these are the belief that gold is G and the belief that fool’s gold is F. Importantly, our
subjects take their beliefs to have a certain logical relation. This is clear because they are
inclined to make a certain inference on the basis of those beliefs—an inference they
would express as ‘gold is F and G.’ Thinking out loud, they might say, “gold is F; gold is
G; therefore, gold is F and G.” If externalism is right, however, the beliefs in question do
not have that logical relation. That is, according to externalism, no such inference is
warranted. It would be fallacious to make any deductive inference on the basis of those two beliefs. The beliefs are simply about different things.9

To make the general point more explicit: all my objection requires (and what is clearly possible) is a situation in which an agent uses a word with (what we might tendentiously call) a single sense to express what, according to externalism, must be regarded as two logically independent meanings. It will then turn out that (given certain other plausible assumptions) we must attribute to the agent two logically independent beliefs. These beliefs will be such, however, that the agent will be disposed to draw a deductive inference from them and will be in principle unable to avoid what will amount to an invalid deductive inference.

Now, any view can be maintained, come what may, by swallowing its unpalatable consequences. Some externalists might take my argument not as a refutation but as an elucidation of an interesting epistemic consequence of externalism. It turns out to be false, according to this strategy, that empirical ignorance should be necessarily insufficient for incoherence. Sometimes, if conditions conspire, a little bit of ignorance can render you logically incoherent and in principle unable to detect and correct your failing. Such a strategy can only be considered a desperate manoeuvre. Our conviction that logical coherence is necessarily independent of empirical knowledge is stronger than the controversial analysis that grounds externalism.

9 The example could also be of an agent who is a member of two linguistic communities differing only in that one uses ‘gold’ for gold while the other community uses that term for fool’s gold. Unaware of this, our agent, when in one community, says, ‘gold is G,’ thereby expressing, according to externalism, her belief that gold is G. Later, in the other community, she says, ‘gold is F’ and thereby expresses, according to externalism, her belief that fool’s gold is F. She would be inclined, however, to draw an invalid deductive inference from those beliefs. Cf. Loar (1988, pp. 106–7).
Notice how Kripke’s puzzle about belief is importantly different from that presented here. In his puzzle one uses a word with (what I will tendentiously say are at least) two senses to express what, according to Millianism, must be regarded as a single meaning. It then turns out that (given certain other plausible assumptions) we must attribute logically contradictory beliefs to an agent who will be unable in principle to detect or avoid them. Thus, in Kripke’s puzzle, from the point of view of the anti-Millian, the crucial element is that a word, ‘London’ or ‘Paderewski,’ which has two senses, is treated as expressing a single definite meaning (because the word’s reference is unambiguous). By contrast, in my example, from the point of view of the internalist, the crucial element is that a word, ‘gold,’ which has a single sense, is treated as expressing two meanings. This difference is crucial: the problem with Millianism, from the point of view of anti-Millianism, is that it individuates meanings too coarsely. The problem with externalism, from the point of view of internalism, is that it individuates meanings too finely.

It is easy to imagine the agents in my example inferring, from their beliefs, another belief, a belief they would express as ‘gold is F and G.’ In making this inference, do the agents

(i) believe that gold is F and G but fail to believe that fool’s gold is F and G;
(ii) believe that fool’s gold is F and G but fail to believe that gold is F and G;
(iii) believe that gold is F and G and also believe that fool’s gold is F and G; or
(iv) fail to believe that gold is F and G and also fail to believe that fool’s gold is F and G?

10 In Kripke’s puzzle, it is the agent attributing the beliefs that is ultimately the relevant language user—rather than the agent attributed the beliefs. But this distinction makes no significant difference.
Neither (i) nor (ii) is very promising. First, there seems to be no reason in principle for preferring (i) to (ii) or vice versa. A connected problem: with each of (i) and (ii) we attribute to the agents one deductive inference while denying them another; but it is unclear why, given the example as rendered above, the agents would make only one of the inferences. They are in analogous positions with respect to each of the deductive inferences. Given the example as rendered, what justifies attributing to them one of the inferences while denying them the other?

More importantly for my purposes, (i) and (ii) share an important disadvantage. With either of them, we attribute to the agent a fallacious deductive inference that he is in principle not in a position to avoid. That is, either of (i) or (ii) has the consequence that the agents are epistemically faulty, in the grave manner of logical incoherence, simply by reason of ignorance. They are inclined to make invalid inferences and no amount of logical acumen will eliminate this inclination—what they need to discover is the chemical fact that the stuff in the mine has a different composition from the stuff on the surface of the earth. With either of (i) or (ii) the agents in question will have made (what they would characterize as) a deductive inference on the basis of beliefs that have no special logical relation to each other. To make a deductive inference from beliefs that do not have the proper logical relation represents a serious epistemic failing. It indicates that you have not maintained the proper logical order among your beliefs. The problem is that if externalism is true, a little bit of ignorance will suffice to put you into this state. So neither of (i) or (ii) is acceptable.

Now, with (iii) we attribute to these agents two fallacious deductive inferences. There are a number of problems with this alternative as well. First, the agents would claim to be making a single inference where we would attribute two inferences. Moreover, and more importantly, we exacerbate the more serious worry attending each of the earlier options. With (iii) the agents in question are twice as inclined to make invalid
deductive inferences as they were under (i) or (ii). Neither the belief that gold is F and G nor the belief that fool’s gold is F and G follows from the agents’ beliefs. If we accept (iii), however, the agents make both of these invalid deductive inferences from their earlier beliefs.

So (iii) is unacceptable as well. What about (iv)? With (iv) we deny the agents the only plausible candidates for the conclusion of their inference. If the belief they infer from their antecedent beliefs is neither that gold is F and G nor that fool’s gold is F and G, then it is unclear what belief they are inferring from their antecedent beliefs. They are making some deductive inference or other (which *ex hypothesi* is a deductive inference from their antecedent beliefs) and assuming they are rational, their inference should be valid—according to externalism, however, no such valid inference seems forthcoming.

None of (i) through (iv) is acceptable; but given the *externalist characterizations* of the mental states of the agents in the example, they seem to exhaust the alternatives. Of course, if we interpret the agents as believing that, say, substances with certain macro properties are G and as believing that substances with those macro properties are F, then an attractive characterization of their inference is as to the conclusion that substances with those macro properties are F and G. My purpose, however, is not to argue in favor of internalism except by presenting a problem for externalism. If externalism is correct and the agents’ beliefs are that gold is G and that fool’s gold is F, then there seems no acceptable way to characterize the deductive inference that those agents will be inclined to make on the basis of those antecedent beliefs.

There is a significant corollary to this example. We can imagine a different agent who came to the belief, as he would put it, that gold is F and G *not* (invalidly) on the basis of his beliefs that gold is G and that fool’s gold is F, but on the basis of his beliefs that gold is G and that *gold* is F. He might have formed the belief that he would express
as “gold is F” not as the result of interaction exclusively with fool’s gold but as the result of interactions exclusively with gold. That is, this new agent might only have had interactions with gold, never with fool’s gold, and might have come to the beliefs he would express as “gold is F” and “gold is G.”

Why is this a significant corollary? Because this agent might nevertheless be internally just like the agents above. It is reasonable to suppose that externalism would characterize this agent’s beliefs as that gold is F and that gold is G. So when this agent concludes that gold is F and G, he is making no invalid inference. He has maintained logical coherence, drawing an inference only when it is valid. But he might be an internal duplicate an agent in the example above. We would then have internal duplicates one of whom is characterized as rational, the other irrational.

This example supports my claim that externalism undermines our view of logical coherence as a minimal condition on rationality. By making our internal state insufficient for the individuation of our mental states, externalism has the following consequence: we may be inclined to draw inferences based on perceived relations between two beliefs when the beliefs do not in fact satisfy the perceived relations. These inferences will be invalid but we will be in principle unable to avoid making them. If our mental states are beyond us, then we cannot be held responsible for maintaining proper logical order among them.

Paul Boghossian (1992, 1994) argues along related lines. I have a different view from his of the relation between Kripke’s puzzle and the kind of example presented here, as well as a different view of the relation between Millianism and externalism. The kind of example he uses involves an agent’s “switching” from Earth to Twin-Earth, which, together with an essential temporal element, complicates intuitions. In addition, his argument requires the controversial premise that the expression of a memory belief should not vary in reference, no matter how long the agent is on Twin-Earth, even if the expression of that memory belief involves a natural kind term. This premise is rejected by Ludlow (1995).
Furthermore and perhaps more importantly, externalism will not allow us to see logical coherence or, since logical coherence is a minimal condition on rationality, rationality as determined by overall internal state. Coherence is not basically a matter of “having one’s house in order.” Duplicates can differ with respect to coherence and rationality. These are grave consequences.
Chapter 3: Russell and Rigidity

In *Naming and Necessity* Saul Kripke argues for a series of interconnected conclusions about the semantics of certain kinds of terms and about the modal and epistemic status of sentences involving those terms. Many take him to have displaced the dominant so-called “Frege–Russellian” theory of names. Kripke uses a variety of arguments against the theories he wants to reject. Here I will focus on one of those arguments.

In the preface Kripke briefly describes the genesis of some of the leading ideas of his monograph. In particular he gives some further explanation of certain points related to modality and rigid designation. Of course a thorough treatment of these points cannot be expected within the bounds of a preface. His discussion is meant only to amplify, in a limited way, the main text. Still, in that preface, Kripke gives, in broad outlines but perhaps more clearly than anywhere else in the text, an important argument against the Russellian thesis that names are disguised (or abbreviated) definite descriptions. It is this argument, the so-called modal argument, and Kripke’s reply to a possible objection to the argument, that are my focus here.

Why focus so narrowly on one argument by one philosopher against the view of another? Kripke’s argument—though, by his own admission, not stated thoroughly or in detail—has been widely influential. Many today take Kripke (together with Donnellan and Putnam and others) to have decisively refuted the Russellian view of names. The argument on which I will focus may seem to be the most decisive. It is interesting too because of its connection with Kripke’s thesis of names as so-called “rigid designators”—a thesis for which he is responsible and celebrated. Also, I believe that Russell’s view contains deep insights into the nature of the referential or intentional relation, insights which are less well appreciated today. Only by dispelling the
widespread feeling that Russell’s theory is inadequate to actual linguistic data can we begin to understand again some of Russell’s deeper philosophical motivation.

If I am right, the distinction Kripke wishes to draw between names and descriptions can be understood in terms of a distinction between two ways of understanding operators (de re and de dicto), a distinction that applies equally whether a name or a description appears in the context of the operator. I do not pretend to be the first to attempt a defense of Russell along these lines.\(^1\) However, previous defenses may have suffered from flaws that interfered with their effect. I hope to present a new defense, one that takes full account of Kripke’s challenge.

The basic idea

The basic idea of Kripke’s argument is that names are rigid designators and definite descriptions are not. Hence names cannot be simply disguised definite descriptions. A symbol is a rigid designator, according to Kripke, if in every possible world it designates the same object (1980, p. 48). Compare the following:

\[
\begin{align*}
(1) & \quad \text{Aristotle was fond of dogs.} \\
(2) & \quad \text{The last great philosopher of antiquity was fond of dogs.} \\
(3) & \quad \text{Exactly one person was last among the great philosophers of antiquity, and any such person was fond of dogs.}
\end{align*}
\]

According to Kripke, Russell thinks that (1) should be analyzed as something like (2), and that (2) should in turn be analyzed as (3) (pp. 6–7). Kripke will reject this two-part analysis.

“[E]veryone agrees that there is a certain man—the philosopher we call ‘Aristotle’—such that, as a matter of fact, (1) is true if and only if he was fond of dogs” (1980, p. 6). Kripke’s point is that, ignoring the counterfactual situations in which Aristotle would not have existed, the truth conditions of (1) as it describes counterfactual situations are the same as in the actual world. Since the truth conditions of (1) with respect to counterfactual situations involve the same man involved in (1)’s actual truth conditions—Aristotle—‘Aristotle’ in (1) must designate the same object in those counterfactual situations as it does in the actual world (ignore the situations in which Aristotle would not have existed). Thus, ‘Aristotle’ in (1) is a rigid designator. And the argument generalizes to all proper names.

Except for a technical detail to be discussed below, Russell (at least on the extension of his view that I will propose below) can largely accept this conclusion.² My criticism of Kripke’s argument is not primarily that he does not establish his first premise—that proper names are rigid designators. I might limit its generality (perhaps

² I do not accept it in full generality; that is, I believe there are sentences in which proper names are used non-rigidly. There was a time when the truth conditions of “Neptune has no moons” as it described counterfactual situations may not have involved the same object as its actual truth conditions. If a scientist had said “Neptune has no moons” a particular planet would have been relevant to the actual truth conditions of that sentence. But the truth conditions of that same sentence with respect to counterfactual situations would have involved a different planet (perhaps even something other than a planet, some cosmic animal). It seems to me, in other words, that there was a time when ‘Neptune’ was clearly a nonrigid name. Though it’s less clear to me that it is nonrigid now, I’m sure there are now nonrigid names on the same model. Kripke has a different intuition about ‘Neptune’ (see 1980, pp. 79n33, 96n42). Thanks to M. Johnston for pointing out that in the context of the dispute over the authorship of certain distinguished plays, the name ‘Shakespeare’ appears to be used non-rigidly to denote whoever wrote them.
not all names are rigid designators), but Kripke might not reject that limitation, and it is in any case independent of my main objection. The next step of the argument, however, is more problematic.

Kripke claims that to think that (1) is (an abbreviation of (2) and thus is) analyzed as (3) will conflict with what he has said about the rigidity of proper names. I reject this claim. I believe that we can see (1) as analyzed by (3) and accept that proper names are usually rigid designators.

It is worth considering certain related issues before proceeding to the central argument. At one point Kripke makes the following assertion, “I assume that Russell is right in that definite descriptions can at least sometimes be interpreted nonrigidly” and later, “my principal thesis…contrasts names with nonrigid definite descriptions, as advocated by Russell” (1980, p. 6n8). Further below: “I am relying on the fact that in practice Russell invariably interprets ordinary names nonrigidly” (1980, p. 10, no citation to Russell). Again: “Frege and Russell certainly seem to have the full blown theory according to which a proper name is not a rigid designator” (1980, p. 58, no citation to Frege or Russell). Elsewhere, however, he asserts that Russell “did not consider modal questions; and the question of the rigidity of names in natural language was rarely explicitly considered after him” (1980, p. 14).

Kripke cannot mean that Russell claimed explicitly that descriptions can be interpreted nonrigidly and that names are nonrigid designators. Rather, Kripke is saying that it follows from Russell’s theory that names and the definite descriptions they abbreviate are, at least sometimes, nonrigid. What is important for Kripke’s position is that the definite descriptions Russell saw names as abbreviating be nonrigid. Kripke reads Russell as advocating a view that implies that the definite descriptions abbreviated
by names are nonrigid. But I will argue that Russell’s theory need not have that consequence and that it can accommodate all of Kripke’s data.

One of the most important aspects of Russell’s theory is that definite descriptions and the names that abbreviate them are not really terms. The surface grammar of sentences containing definite descriptions or ordinary (as opposed to logically proper) names is misleading, according to the theory. All such sentences actually harbor a quantificational structure and express general, rather than singular, propositions. We should not ask for the meaning of a definite description or a name taken in isolation, because it has no meaning in isolation. It is an incomplete symbol, requiring a context in order to contribute to the meaning of the more complex expression of which it will then form a part. So, technically, Russell has to hold that neither names nor definite descriptions, per se, designate the same object in every possible world. According to Russell, ordinary names and definite descriptions do not designate any object in any possible world. Rather, names and definite descriptions combine with the rest of the sentence in which they appear to express general propositions which can then be satisfied by objects.

---

3 By ‘Russell’s theory’ I mean what has come to be widely regarded as such, which is why I can avoid detailed documentation—and Kripke may have had the same justification.

4 Some may take this consequence to be sufficient grounds for rejecting Russell’s theory. It may be implausible to suppose that expressions like ‘the last great philosopher of antiquity’ or ‘Aristotle’ are not terms and do not have a meaning in isolation. For the purposes of discussing Kripke’s argument, however, I will assume that he is willing to allow Russell this thesis. Kripke cannot assume, for example, without begging the question against Russell, that ordinary names are terms with a complete meaning even in isolation. (It does not seem to me that he does make this assumption.) Otherwise the argument that names are not disguised definite descriptions would be too easy; definite descriptions (according to Russell’s own theory) are not terms that stand for individuals; but the names Russell claims abbreviate those descriptions clearly are terms that stand for individuals. Therefore, names are not disguised definite descriptions.
A better formulation of the thesis of rigid designation would be that the truth conditions for a sentence containing a proper name do not vary even when the sentence is evaluated with respect to counterfactual situations. This formulation preserves the philosophical content of the original formulation—the truth conditions, with respect to counterfactual situations, of a sentence containing a proper name involve the same individual that is involved in the actual truth conditions of that sentence. But the new formulation does not stipulatively settle the issue whether Russell’s theory is incompatible with the thesis of rigid designation. If the truth conditions for a sentence containing a definite description do vary when the sentence is evaluated with respect to counterfactual situations, then Russell’s theory that names are simply disguised definite descriptions is in trouble. But if they do not, the theory may resist Kripke’s objections.

It seems to Kripke that viewing (1) as analyzed by (3) will conflict with viewing ordinary proper names as rigid designators. Why does he believe there is this conflict?

The actual truth conditions of (3) agree extensionally with those mentioned above for (1), assuming that Aristotle was the last great philosopher of antiquity. But counterfactually, Russell’s conditions can vary wildly from those supposed by the rigidity thesis. With respect to a counterfactual situation where someone other than Aristotle would have been the last great philosopher of antiquity, Russell’s criterion would make that other person’s fondness for dogs the relevant issue for the correctness of (1)! (1980, p. 7)

I will argue that this underappreciates the resources of Russell’s theory. I do not think that the truth conditions predicted by Russell’s theory for (1) with respect to a counterfactual situation where someone other than Aristotle would have been the last great philosopher of antiquity, need necessarily involve that other person’s fondness of dogs. In fact I do not think that Russell’s theory clearly establishes a single set of truth conditions, with respect to a counterfactual situation, for any simple sentence containing a name or a definite description.
Russell’s theory of descriptions predicts that sentences containing names or definite descriptions in the context of any operator will be ambiguous. There will be at least two ways to represent the logical form of such sentences consistent with the contextual definitions for definite descriptions Russell has given. Therefore there will be no single set of truth conditions for sentences containing names or definite descriptions in the context of an operator. Of course none of (1), (2), or (3) has any operators (if we take the copulas to be tenseless, see p. 11n14). Russell’s theory does not predict any ambiguity with respect to their actual truth conditions. I would claim, however, that the situation becomes more complicated when we consider counterfactual truth conditions.

The proposal I would defend

Here, informally, is the proposal I would defend. The truth conditions for (1) with respect to a counterfactual situation, say S, are given by:

(1’) With respect to S, Aristotle was fond of dogs.

That sentence is a complex sentence in which a simple sentence is governed by the (modal) operator ‘with respect to S’. Russell’s theory predicts that the complex sentence is ambiguous, depending on whether the description that replaces the embedded name (‘Aristotle’) has a primary or a secondary occurrence, that is, depending on whether the quantifier that is introduced when the description is eliminated includes the operator in its scope, or on the contrary is included in the scope of the operator. Supposing, for convenience, that ‘Aristotle’ abbreviates ‘the last great philosopher of antiquity’, the sentence, “with respect to S, Aristotle was fond of dogs” can mean either (i) that exactly one person was last among the great philosophers of antiquity, and any such person was, with respect to S, fond of dogs, or (ii) that with respect to S, exactly one person was last among the great philosophers of antiquity, and any such person was fond of dogs (where
(i) and (ii) are themselves evaluated at the actual world. (i) corresponds to treating the
description to replace ‘Aristotle’ in (1) (and so derivatively, let us say, to treating the
name itself) as having wide scope; (ii) corresponds to treating it as having narrow scope.

If we were to treat the name as having wide scope (so to speak), then the truth
conditions for (1') would be the same as those Kripke feels intuitively are appropriate to
(1) when it describes a counterfactual situation S. If I am correct that (1') gives the truth
conditions for (1) with respect to any counterfactual situation S, then Russell’s theory is
consistent with Kripke’s intuitions about the rigidity of ordinary proper names. The
proposal can be put another way: when Kripke says, “Russell’s criterion would make that
other person’s fondness for dogs the relevant issue for the correctness of (1),” he is
illegitimately assuming a narrow scope reading of (1').

Kripke wants to show that (2) cannot abbreviate (1). If it did, Kripke argues, then
(3) would analyze (1)—granting that (3) analyzes (2). Analyses must establish necessary
equivalence. There can be no world in which analysandum is true while analysans is
false (or vice versa). But, urges Kripke, it is clear that there are worlds in which (1) is
ture while (3) is false. Imagine a world in which Aristotle is not the last great philosopher
of antiquity (a circumstance that does not dampen his fondness of dogs), a world,
however, in which the last great philosopher of antiquity is himself not fond of dogs. That
would be a world in which (3) is false but (1) is true, says Kripke.

Let us call the world in question ‘S’. Now, my response proceeds, Kripke’s point
is made only if S is a world in which (3) is false but (1) is true. (1) is true with respect to
S only if (1') is true. (1'), however, is ambiguous. It is true; but only on one interpretation.
Thus, what Kripke says is true only on one interpretation. Perhaps that is the most natural
interpretation. But it need not be seen as inconsistent with Russell’s theory that (1) is true
with respect to S; indeed, according to my proposal, it can be seen as predicted by the resulting theory.

Focus for a moment on the following sentence: (2') In S, the last great philosopher of antiquity was fond of dogs. I claim that the ambiguity predicted by Russell’s theory for (2') is reflected in an intuitive ambiguity in our understanding of that sentence. Still, (2) is analyzed by (3). So there can be no possible world in which (2) and (3) have different truth values. Now consider the following sentence: (3') In S, exactly one person was last among the great philosophers of antiquity, and any such person was fond of dogs. If (3') is unambiguous, but (2') is ambiguous, then there may be an interpretation under which (2') and (3') have different truth values. The crucial consequence is that the equivalence between (2') and (3') does not hold unambiguously. Furthermore, though (2) and (3) are each unambiguous and logically equivalent, expressions of that equivalence may themselves be only ambiguously true. “For all worlds w, (2) is true in w if and only if (3) is true in w,” for example, will be only ambiguously true.

The moral is important. Even if two sentences \( p' \) and \( q' \) are each unambiguous and necessarily equivalent, it does not follow that the sentences \( O p' \) and \( O q' \) (where O is some modal operator) must have the same truth value. Two otherwise identical sentences, differing only in the substitution of an expression in one by a necessarily equivalent expression in the other sentence, can have different truth values as a result of an ambiguity differentially affecting (which may be a matter of mere convention) one of the sentences.

I review this in order to show that we cannot conclude from the fact that there is an S for which (1') is true but (3') is false that (1) and (3) themselves are not necessarily equivalent. It might be that the necessary equivalence between (1) and (3) is expressed by a sentence that is only ambiguously true. Kripke does show that there is an S for which
(1’) is true but (3’) is false, but (1) and (3) may still be necessarily equivalent (under the relevant disambiguation). Kripke would have to show further, what cannot be shown, that the reading of (1’) under which it is true corresponds to the narrow scope reading of (1’). If that were so, if the narrow scope interpretation of (1’) were true, then, in the sense Kripke needs, there would be a possible world in which (1) was true but (3) was false. Otherwise he has shown only that there is an interpretation of “Necessarily, (1) if and only if (3)” under which it is false. That may be so; but does not itself disprove Russell’s theory. Again, according to my proposal, it is what we should expect.

In any case, if the mere fact that there is a counterfactual situation S relative to which (1) is true but (3) is false showed that (1) and (3) are not necessarily equivalent, then by parallel reasoning it could be shown directly not only that (1) and (2) are not necessarily equivalent but also that (2) and (3) are not necessarily equivalent.

After all, there are worlds in which the last great philosopher of antiquity is neither a philosopher nor fond of dogs. These are worlds in which (2) is false. In some of those worlds (3) will still be true.

In one good sense, to take a world in which the last great philosopher of antiquity is not fond of dogs is just to take a world in which Aristotle is not fond of dogs: the last great philosopher of antiquity is Aristotle and so to take a world in which that philosopher is not fond of dogs is to take a world in which Aristotle is not fond of dogs. But of course that world might have a different last great philosopher of antiquity; and that other philosopher might be fond of dogs. Thus, I claim there is a good sense in which it is possible for (2) to be false while (3) is true.

But Kripke’s rejection of Russell’s two-part analysis is meant to eliminate only the equivalence between (1) and (2)—and not the equivalence between (2) and (3). Indeed, Kripke needs the equivalence between (2) and (3), together with the
inequivalence between (1) and (3), to prove the inequivalence between (1) and (2). If I’m correct, however, we can as easily (or hardly) show that (2) and (3) are not equivalent as that (1) and (3) are not equivalent.

I have been claiming that Kripke’s argument presupposes the validity of the following schema:

\[(Z) \quad (p+q) \not\in \square (\exists S, p' \& \neg \exists S, q')\]

Kripke wants to argue by contraposition against the equivalence of (1) and (2)—letting (1) and (2) be values for ‘p’ and ‘q’. But there is good reason (viz. possible ambiguity) to doubt the validity of (Z). In any case, Russell’s theory could, given one interpretation, predict counterexamples (cf. the example of (2) and (3) above) to the schema. Without (Z), however, Kripke’s argument is insufficient to reject Russell’s theory.

Elaboration of the proposal

In this section I will address a number of concerns that may have arisen with respect to my proposal. The first concern is whether the proposal is one Russell himself could have made in response to Kripke’s objection.

Here’s an argument that he could not have made this proposal. Russell’s ‘contextual definition’ of ‘the F is G’ is that there exists an individual who is, uniquely, F, and that individual is G. Russell meant that for any circumstance in which ‘the F is G’ is true, ‘there exists an individual who is, uniquely, F, and that individual is G’ is also true (and vice versa). Therefore, the argument proceeds, Russell held:

\footnote{Thanks to S. Soames for comments and discussion that prompted these elaborations.}
(R) for all possible worlds w, ‘the F is G’ is true in w if and only if ‘there exists an individual who is, uniquely, F, and that individual is G’ is true in w.

Now, if, furthermore, Russell took that biconditional to be, itself, unambiguously true, then, perhaps, my proposal would conflict, to that extent, with Russell’s view.

I have claimed that ‘with respect to W, the F is G’ is ambiguous; and I have allowed that ‘with respect to W, there exists an individual who is, uniquely, F, and that individual is G’ might be unambiguous. But if ‘‘there exists an individual who is, uniquely, F, and that individual is G’ is true in w’’ is supposed to have exactly the same interpretation as ‘with respect to w, there exists an individual who is, uniquely, F, and that individual is G’ (so that the former is ambiguous if and only if the latter is), and if ‘‘the F is G’ is true in w’’ has exactly the same interpretation as ‘with respect to w, the F is G’ (so that the former is ambiguous if and only if the latter is), then (R) is not unambiguously true. ‘‘The F is G’ is true in w’ will come out ambiguous; but ‘‘there exists an individual who is, uniquely, F, and that individual is G’ is true in w’’ will come out ambiguous. Russell’s biconditional will come out true only on one disambiguation of ‘‘the F is G’ is true in w.’’

This argument does seem to me to show that it follows from my proposal that (R) will be ambiguous. But as Kripke himself has noted, Russell did not consider modal questions. I am not sure whether Russell himself would have been averse to the idea that the equivalence he was claiming would lead to an ambiguity in (R). Admitted, (what we might call) a wide scope reading of ‘‘The F is G’ is true in w’’ would make the biconditional false; while that is the scope we need to accommodate the rigidity intuition. But I cannot see that this fact is sufficient to undermine the adequacy of Russell’s analysis. It might just show the inadequacy of (R) for representing Russell’s analysis. Again, both scopes are available for ‘‘with respect to w, the F is G.’’
Consider the following parody of my proposal: ‘the F is G’ is not true just in case

(i) it is not true that the F is G.

However, (it is not true that the F is G) if and only if either

(ii) there is no unique F that is G (which is true if there’s no F)

\textit{or}

(iii) some F is not G (which is false if there’s no F).

So (i) is ambiguous (the ‘not’ can amount to sentence-negation or predicate-negation, so to speak). But \textit{according to the kind of argument I have been urging, I would now have to say that “‘the F is G’ is not true” is also ambiguous.} This ambiguity would not be predicted by Russell. The mechanism by which scope ambiguities arise in Russell’s theory of descriptions simply does not apply in this case (because the description is within quotation marks). How can I respond to this challenge?

I would be happy to be understood as proposing a friendly \textit{addition} to Russell’s theory. The ambiguity of “‘the F is G’ is not true” arises, on my proposal, as a result of (a) the ambiguity of ‘it is not true that the F is G’ (which \textit{is} predicted by Russell’s theory), and (b) the equivalence of “‘the F is G’ is not true” and ‘it is not true that the F is G’ (which I claim is independently plausible). \textit{Mutatis mutandis} for my proposal above.

But now consider the following:

\begin{align*}
\text{(m) } & \text{‘the F is G’ is true if and only if there is a unique F} \\
& \text{and whatever is F is G.}
\end{align*}

This is surely \textit{un} ambiguously true. It is just a statement of Russell’s contextual definition. But if (1) holds, then the following must hold.

\begin{align*}
\text{(n) } & \text{‘the F is G’ is } \textit{not} \text{ true if and only if it is } \textit{not} \text{ the case} \\
& \text{that there is a unique F and whatever is F is G.}
\end{align*}
I have claimed that sentences like (n) are ambiguous and have a reading under which they are false (*i.e.*, the wide scope reading, so to speak, of the left-hand side). How can this be?

The crucial point is that although (n) follows from (m), that (m) is unambiguous does *not* entail that (n) is unambiguous. (n), like (m) is true. But unlike (m), (n) has an interpretation under which it is false. Introducing a negation operator on both sides of an equivalence cannot change the truth value of the biconditional; *but it can change certain other semantic properties of that biconditional*. In particular, the biconditional can become ambiguous. Another way to put it is that although (n) follows from (m), that implication is not itself unambiguous.

Of course, the ambiguity I’m urging is not easily represented in our standard formal apparatus. It corresponds to treating “‘the F is G’ is not true” as having a reading under which it is equivalent to “‘the F is not G’ is true” (and to interpreting the description in ‘the F is not G’ as having wide scope); but how to represent moving the ‘not’ within the quotation marks? Formally, we treat items within quotation marks as a single syntactic unit. Nevertheless, the equivalence seems intuitively to hold (on one disambiguation of ‘it is not true that the F is G’).

Similarly, returning now to my proposal (and to modal operators rather than the extensional negation operator), I urge that

\[(p) \quad \text{for all worlds } w, \text{ ‘the F is G’ is true with respect to } w \text{ if and only if with respect to } w, \text{ the F is G} \]

has an interpretation under which it is equivalent to the following.

\[(q) \quad \text{for all worlds } w, \text{ ‘the F is G’ is true with respect to } w \text{ if and only if there is a unique F and, with respect to } w, \text{ it is G}. \]
Standardly, modal operators are understood ‘world-neutrally.’ This amounts to adding a world parameter to the predicates that occur in the scope of the modal operator. Thus, standardly, ‘with respect to w, the F is G’ would be understood as ‘the unique F in w is G in w.’ But this is just to assert that in modal logic, standardly, modal operators are understood to take wide scope over the definite descriptions that might occur within their scope. On the other hand, it should be clear that according to Russell, ‘with respect to w, the F is G’ is ambiguous. There is what amounts to a ‘world-partial’ interpretation of that sentence—an interpretation in which ‘the F’ is not given a world parameter. It is to this fact that I am implicitly appealing in urging (q) as one possible interpretation of (p).

A response to my proposal

Another response (although one closely related to the objections discussed above) to my proposal may be immediately appealing: If indeed (1’) gives the truth conditions for (1) with respect to a counterfactual situation S, then since the truth conditions for (1) with respect to a counterfactual situation S will be not be given by the wide scope reading of (1’), the narrow scope reading of (1’) is required—eliminating the Russellian rebuttal of Kripke’s objection. But why are the truth conditions of (1) with respect to a counterfactual situation S not given by (1’) when the name is interpreted as having wide scope?

The following thesis might seem to have that consequence:

(T) (3) is true in a counterfactual situation S iff exactly one person is last among the great philosophers of antiquity in S and any such person is fond of dogs in S.

Here’s why (T) might appear to have the claimed consequence. Suppose, as per (T), that the truth conditions for (3) with respect to a counterfactual situation S depend on whether
exactly one person is last among the great philosophers of antiquity in S, and on whether any such person is fond of dogs in S. Then, if we analyze (1) as equivalent to (3), the truth conditions of (1) with respect to a counterfactual situation S must depend on those things too.6 But the truth of (1'), if we give the name a wide scope reading, does not depend on the fondness of dogs of the last great philosopher of antiquity in S. Therefore, (1') with a wide scope reading does not give the truth conditions of (1) as it describes the counterfactual situation. Thus, if the truth conditions for (1) with respect to a counterfactual situation S are given at all by (1'), the name in (1') must be given a narrow construal.

The narrow construal of the name in (1') makes its truth conditions involve the fondness for dogs (in S) of the last great philosopher of antiquity in S. The claim that ‘Aristotle’ in (1) is rigid amounts precisely to the claim that its truth conditions depend only on Aristotle’s fondness for dogs. And we can suppose that in S, Aristotle is not the last great philosopher of antiquity. The proposed accommodation for Kripke’s intuitions would thus fail. So this response to my proposal could be summarized as follows: (i) (T) is obviously true, and (ii) (T) implies that the name in (1') must be given a narrow construal. This would make it impossible to accommodate the rigidity intuition in the way I have proposed.

I believe this objection to my proposal is unsuccessful. It is not easy to see why, however.7 (T) is, I believe, true. And the truth conditions for (1) as it describes a counterfactual situation are the same as the actual truth conditions for (1'). But I do not believe that for this to be true the name in (1') must be given a narrow scope reading. So I deny (ii). Isn’t this position inconsistent? I do not think so.

---

6 See next footnote.
7 I have tried to make it easier to see why by italicizing the offending line above. See previous footnote.
My argument depends on the general truth of the following substitution schema (where the substitution class for the substitutional variable ‘R’ contains sentences, and that for ‘S’ contains names of counterfactual situations):

(U) “R” is true in S iff in S, R.

Consider the analog of (T) for (2):

(T2) (2) is true in a counterfactual situation S iff exactly one person is last among the great philosophers of antiquity in S and any such person is fond of dogs in S.

‘(2)’ functions here as an abbreviation of “the last great philosopher of antiquity was fond of dogs”. Therefore, (T2) will be true if and only if the following is true:

(T3) “The last great philosopher of antiquity was fond of dogs” is true in a counterfactual situation S iff exactly one person is last among the great philosophers of antiquity in S and any such person is fond of dogs in S.

By (U), however, we obtain the following:

(T4) “The last great philosopher of antiquity was fond of dogs” is true in a counterfactual situation S iff in S, the last great philosopher of antiquity was fond of dogs.

Notice what follows the “iff” in (T4). According to Russell, “in S, the last great philosopher of antiquity was fond of dogs” is ambiguous. So, according to Russell’s theory, (T4) attributes to what precedes the “iff” different truth conditions depending on how we disambiguate what follows that “iff.”

Again, to some, this result will be problematic. How can “the last great philosopher of antiquity was fond of dogs’ is true in a counterfactual situation S” be
ambiguous? The sentence containing the description is ensconced in quotation marks and is thus protected from any scope interaction with the modal operator. The situation in which scope ambiguity is created in Russell’s theory does not arise here. True enough. But how can we deny that “‘The F is G’ is true in S” is equivalent to “In S, the F is G”?

Whatever rules might be embedded in our formal practices, it should be clear that *intuitively*, to say that “‘The center holds’ might be true” is just to say that the center might hold. Compare: “‘The huge summer has gone by’ is false” and “It’s not the case that the huge summer has gone by.”

(T2) and (T3) are in fact acceptable only on one reading. Their right-hand sides give an unnecessary condition for the truth of their left-hand sides, on the other reading. Look again at the right-hand side of (T2), for example. That amounts to the narrow scope reading of the right-hand side of (T4). But there remains the wide scope reading of the right-hand side of (T4). If the condition given by that interpretation were satisfied, then the left-hand side of (T2), on one reading, would be true. And the right-hand side of (T2) might not be satisfied.

This is an important result. (T2), I have argued, is acceptable only on one disambiguation. On the other hand, (T) seems acceptable on any interpretation (indeed it seems to have only one). But (T) differs from (T2) only in the substitution of ‘(2)’ for ‘(3)’. And ‘(2)’ and ‘(3)’ are (abbreviations of) names of equivalent sentences. By now the moral should be clear: “is true in a counterfactual situation S” creates a context in which names of equivalent sentences are not *necessarily* intersubstitutable *salva veritate.*

---

8 Once more: formally, the left-hand side of (T2) does not allow for more than one reading. But intuitively, I claim, it has two. Quite properly, we eliminate the ordinary ambiguity (which is often erased by pragmatic features anyway) in our formal representation. But Russell could appeal to that ambiguity to resist Kripke’s objections.
What is the consequence of all this for the objection to my proposal? Where did the argument for that objection go wrong? The argument tried to show that given (T), if (1) is equivalent to (3), then the truth conditions for (1) as it describes a counterfactual situation S can be given by (1’) only on a narrow scope reading of the contained name. That claim presupposes that the truth conditions for (3) with respect to a counterfactual situation S cannot be different from the truth conditions for (1) with respect to that counterfactual situation. But the equivalence of (1) and (3) is insufficient for this. It should by now be clear that, according to Russell’s theory, (1) and (3) might be equivalent even though the ambiguous (1’) gives truth conditions such that, first, on a wide scope reading of (1’)—under which it is true—(1) as it describes a counterfactual situation S might have different truth conditions from those of (3) as it describes that counterfactual situation; and second, on a narrow scope reading of (1’)—under which it is false—(1) as it describes a counterfactual situation S might have the same truth conditions as (3) as it describes that counterfactual situation.

To put it another way, (1) and (3) might each be unambiguous. Now consider the following claim: For all worlds w, (1) is true with respect to w if and only if (3) is true with respect to w. If “(1) is true with respect to w” is equivalent to “In w, Aristotle was fond of dogs” (and it is), then “(1) is true with respect to w” is ambiguous. “(3) is true with respect to w” will be equivalent only to one interpretation of “(1) is true with respect to w.” So the equivalence will not be unambiguously true. And all this is consistent with the fact that (1) and (3) are themselves unambiguous. When we say that they are equivalent, we say something true (though only ambiguously so). 9

Unless the name in (1’) intuitively requires a narrow scope construal, then my proposal for a Russellian accommodation of Kripke’s objections is untouched. I claim

9 Compare: when we say that banks are financial institutions, we say something true (though ambiguously so).
that (1'), where the name is taken as having wide scope, *can* give the truth conditions for (1) as it describes a counterfactual situation S. This is consistent with (T). Like (T2) and (T3), the following has a reading under which it is false:

\[(T5) \quad \text{“Aristotle was fond of dogs” is true in a counterfactual situation S iff the last great philosopher of antiquity in S was fond of dogs in S.}\]

I see no reason to insist that we disambiguate (T5) so that it is true; and so long as Russell’s theory can predict a reading under which it is false, I reject the objection to my proposal.

Possible misconceptions

Later in the preface (pp. 10ff) Kripke considers what he calls a ‘misconception’ concerning the relation of rigidity to scope. He disclaims a sufficient discussion of the issue, and admits that relevant considerations in the text seem to have been stated too briefly. Moreover, he acknowledges that it would be possible to handle many of the linguistic intuitions adduced on behalf of rigidity by reading names in various sentences as nonrigid designators with wide scopes, analogously to wide scope descriptions. But he thinks that some linguistic phenomena support the rigidity intuition as opposed to an explanation in terms of scope.

The proposal I defend does not interpret Kripke’s intuitions concerning rigidity as the result of scope ambiguities *instead* of rigidity. I come to praise rigidity, not to bury it. My proposal amounts to accommodating rigidity, not to replacing it. This is important: It is not that ordinary proper names are not rigid. They usually are; and interpreting them as disguised definite descriptions that usually take wide scope in the context of modal operators would have that very consequence. Kripke would disagree.
(1) and (2) are ‘simple’ sentences. Neither contains modal or other operators. No scope distinctions about more complex sentences affects the interpretation of these sentences….My view is that ‘Aristotle’ in (1) is rigid, but ‘the last great philosopher of antiquity’ in (2) is not. No hypothesis about scope conventions for modal contexts expresses this view; it is a doctrine about the truth conditions, with respect to counterfactual situations, of (the propositions expressed by) all sentences, including simple sentences. (1980, pp. 11–2)

Let us review this passage in detail. First, it is true that (1) and (2) are simple sentences. But whether scope distinctions about more complex sentences affects the interpretation of these sentences seems to me to depend on the kind of interpretation in question. Modal interpretation of these simple sentences (i.e., interpretation of these sentences with respect to counterfactual situations) may well, I would argue, be affected by—in the sense of not being logically independent of—scope distinctions concerning more complex sentences. This first point is of course the essence of my position.

Second, I would reiterate that ‘the last great philosopher of antiquity’, according to Russell, has no semantic value in isolation. Technically, as noted above, names and definite descriptions cannot be rigid designators on Russell’s theory (according to Kripke’s definition of ‘rigid designator’). The interesting issue, however, is whether Russell’s theory can accommodate Kripke’s intuitions about the truth conditions of sentences containing names. Hence our revision of the definition of ‘rigid designator’.

The important question then is whether Russell’s theory will yield the intuitively correct truth conditions, in counterfactual situations, of simple sentences. Can a wide scope interpretation of modal contexts produce the right truth conditions, with respect to counterfactual situations, for simple sentences? I hold that it can. Kripke holds that it cannot. He goes on to illuminate his position.
…‘It might have been the case that Aristotle was not a philosopher’ expresses a truth, though ‘It might have been the case that the greatest philosopher of antiquity [the last great philosopher of antiquity?] was not a philosopher’ does not, contrary to Russell’s theory….Now the last quoted sentence would express a truth if the description used were read, contrary to my intent, with wide scope. So perhaps it might be supposed that the problem simply arises from an (unaccountable!) tendency to give ‘Aristotle’ a wide scope reading while the descriptions are given a small scope reading;…My point, however, was that the contrast would hold if all the sentences involved were explicitly construed with small scopes…. (1980, p. 13)

It is unclear to me what Kripke means to refer to by ‘the problem’. Just below this passage, Kripke again speaks of ‘the difficulty’ that the Russellian view must account for (and cannot, according to him). But what problem or difficulty does he have in mind? There are some data, perhaps, that need to be accommodated. In particular, the theory should allow that

(a) It might have been the case that Aristotle was not a philosopher

expresses a truth. And it should do this even while allowing

(b) It might have been the case that the last great philosopher of antiquity was not a philosopher

to express a falsehood. Since the theory holds that “Aristotle” in (a) is simply an abbreviation of “the last great philosopher of antiquity” in (b), and since (a) and (b) are otherwise the same, how could it allow the truth values to differ? This the theory does by postulating an ambiguity. On one reading, i.e. the wide scope reading, they are true. On another, i.e. the narrow scope reading, they are false. By adopting different readings, the Russellian can allow their truth values to differ. And we need not claim that “our language and thought are, somehow, impotent to keep the distinction straight” (p. 13), notwithstanding Kripke’s assertion to the contrary.
Kripke’s professed point, however, is that the contrast between (a) and (b) would hold even if they were explicitly construed with small scopes. This is confused. First, on Russell’s theory (and to assume its negation would beg the question), there is no independent notion of the scope of an ordinary proper name. To think otherwise is to deny Russell’s basic principle (which we are taking for granted for the purposes of this discussion) that names and descriptions do not have a meaning in isolation. Ordinary names are not represented logically by an individual constant—this much is not challenged by Kripke. The small scope construal of (a) is by definition equivalent to the small scope construal of (b). Again, if Kripke is in effect claiming to have the intuition that ordinary proper names can be represented logically by an individual constant, the rejection of Russell’s theory is more straightforward (and less interesting; cf. footnote 4 above).

Second, if what Kripke means is that, intuitively, (a) is true while (b) is false, that just shows that if names abbreviate definite descriptions, those descriptions cannot always take small scope in the context of modal operators. But that is not sufficient to show that names cannot abbreviate definite descriptions. It would need to be shown further that (b) is false on a wide scope construal. This has not and indeed cannot be done; (b) is true on a wide scope construal.

I would stress that any intuitions about the truth values of the logical representation of natural language sentences can always be reinterpreted as intuitions about the proper logical representation of those sentences. If what Kripke claims is that, intuitively, (a) is true-on-a-narrow-scope-reading while (b) is false-on-a narrow-scope-reading, either he has already rejected Russell’s representation of the narrow scope reading of (a) (but how to understand the idea of more than one scope for (a) independently of Russell’s theory?) or his intuition really represents a preference for a wide scope reading of (a). Neither situation is a problem for Russell’s theory.
The central example and the intuitive test

In arguing that Russell’s theory is adequate to the considerations adduced as objections, it is important not to overlook the brief discussion of the question on page 62 and in its accompanying footnote 25 in the main text. There Kripke observes, “the teacher of Alexander might not have taught Alexander (and, in such circumstances, would not have been the teacher of Alexander). On the other hand, it is not true that Aristotle might not have been Aristotle…” (1980, p. 62n25). In the preface Kripke refers to this example often (1980, pp. 10, 11, 13) as demonstrating the distinction between names and definite descriptions. The example should thus be evaluated carefully; I will draw a different conclusion from it than that drawn by Kripke.

(c) The teacher of Alexander might not have taught Alexander.

(d) The teacher of Alexander might not have been the teacher of Alexander.

(e) Aristotle might not have been Aristotle.

Kripke invites us to the following analysis. (c) is clearly true. Hence, (d) is true. But (e) is false. Therefore, (d) and (e) cannot be equivalent. I will not deny (c). And I will not argue that (e) is true. My defense of Russell will be that on his theory, taking names and definite descriptions as having wide scope in the context of a modal operator, it is not the case that (d) is true in just those circumstances in which (c) is true. A brief digression will help to illuminate this point.

Russell is adamant that “Socrates is a man” and “Socrates is human” do not express the same proposition—they have different underlying logical forms.10 Assuming

---

10 Though the sentences may be *extensionally* equivalent. See Russell (1919) in Martinich, p. 214 where he says, “[t]he proposition ‘Socrates is a man’ is no doubt
for convenience that ‘Socrates’ abbreviates ‘the teacher of Plato,’ the first sentence could be understood as asserting roughly that there are individuals $x$ and $y$ such that $x$ teaches Plato and any individual who teaches Plato is identical to $x$, and $y$ is human, and $x$ is identical to $y$; schematically,

\[(f) \exists x \exists y (T_P x & \forall z (T_P z \iff x = z) & H_y & x = y)\]

The second asserts that there is an individual $x$ such that $x$ teaches Plato and any individual who teaches Plato is identical to $x$, and $x$ is human.

\[(g) \exists x (T_P x & \forall y (T_P y \iff x = y) & H_x)\]

An important difference is that the ‘is’ in the first sentence is the so-called “‘is’-of-identity” while the ‘is’ in the second sentence is a copula. I hereby echo Russell’s lamentation of our ambiguity in this regard.

Now examine (c). It should be taken as asserting, roughly, that there is an individual $x$, such that $x$ taught Alexander and any individual who taught Alexander is identical to $x$, and possibly $x$ does not teach Alexander:

\[(c') \exists x (T_A x & \forall y (T_A y \iff x = y) & \neg T_A x)\]

This is clearly true. (d), on the other hand, should be taken as asserting, roughly, that there are individuals $x$ and $y$ such that $x$ teaches Alexander and any individual who

---

**equivalent** to ‘Socrates is human,’ but it is not the very same proposition. The *is* of ‘Socrates is human’ expresses the relation of subject and predicate; the *is* of ‘Socrates is a man’ expresses identity. It is a disgrace to the human race that it has chosen to employ the same word ‘is’ for these two entirely different ideas—a disgrace which a symbolic logical language of course remedies.”

Though, confusingly, Russell (1905, 1919) sometimes treats names like ‘Socrates’ as logically proper names (see especially 1919 in Martinich, pp. 216–7).
teaches Alexander is identical to x, and y teaches Alexander and any individual who teaches Alexander is identical to y, and possibly x is not identical to y:

\[(d') \exists x \exists y (TAx \land \forall i (TAi \equiv x = i) \land TAy \land \forall j (TAj \equiv y = j) \land \neg x = y)\]

If this were true, then there would be individuals x and y which though identical are possibly not identical. And Kripke would never accept such a consequence (see p. 3). So (d) will be false if the description is taken as having wide scope in both of its occurrences. If we had the convention that names have wide scope in the context of an operator, then (d') would represent the underlying logical form of (e) as well. This would support an equivalence between (d) and (e).

Notice that if we give either one of the occurrences (but not both) of the definite descriptions in (d) narrow scope, the sentence as a whole comes out true. That is, there is an individual x such that x teaches Alexander and any individual who teaches Alexander is identical to x, and possibly there is an individual y such that y teaches Alexander and any individual who teaches Alexander is identical to y, and (still under the scope of ‘possibly’) x is not identical to y.

\[(d'') \exists x (TAx \land \forall i (TAi \equiv x = i) \land \exists y (TAy \land \forall j (TAj \equiv y = j) \land \neg x = y))\]

Indeed, if we give the description in (c) narrow scope, it too will come out false. It would then be taken as asserting that in some possible world, there is an individual x such that x teaches Alexander, and any individual who teaches Alexander is identical to x, and x does not teach Alexander.

\[(c'') \Box (\exists x (TAx \land \forall i (TAi \equiv x = i) \land \neg TAx))\]

Of course, there is no such possible world.
Finally, if we give both definite descriptions in (d) narrow scope, it will again come out false. There is of course no possible world in which two things that are each uniquely the teacher of Alexander in that world are not identical to each other.

\[(d'') \quad \Box(\exists x(TAx & \forall i(TAi \Box x=i) & \exists y(TAy & \forall j(TAj \Box y=j) & \sim x=y)))\]

The confusion, in moving from (c) to (d) as Kripke does, involves disregarding the effect of the second definite description in (d). In (c) there is only one definite description to expand—in (d), two. If we give the first definite description (the one in subject place) in both (c) and (d) wide scope, then (so long as we take the second definite description in (d) to have narrow scope) both sentences come out true. Similarly, if we take that first definite description to have narrow scope, then (again, so long as we take the second definite description in (d) to have narrow scope) both sentences come out false.

But to conclude that (c) implies (d) is to disregard the differential effect of the second definite description in (d). Unlike the predicate in (c), the definite description in (d), on Russell’s theory, will be replaced by a quantifier phrase whose scope can include or alternatively be included by the modal operator—hence Russell’s emphasis on the distinction between the copula and the ‘is’-of-identity.12

---

12 Moreover, when there is more than one name or description in a sentence, it is important to ‘keep track’ of the relevant scopes. Consider the following sentence (suggested by M. Johnston): Aristotle might have been intermediate in size between Aristotle and Aristotle. In one very clear sense the sentence must be false; and giving all three names wide scope would produce a false formula, as desired. But I have argued that Russell’s theory is sufficiently flexible that other formulae are possible interpretations. Some of these will be true. Is there any sense of the original sentence in which it can be true? If Aristotle might have been much taller than he actually is, and if Aristotle might have been much shorter than he actually is, then there are two ways in which that original sentence might be true: (i) Aristotle might have been [taller than Aristotle-as-he-actually-is and shorter than Aristotle-as-he-might-have-been (had he been much taller than he actually is)] and (ii) Aristotle might have been [taller than Aristotle-as-he-might-have-been (had he
Moreover, definite descriptions can satisfy the intuitive test for rigidity that Kripke gives on pp. 48–9. The test in question is never stated explicitly; but Kripke’s comments suggest the following substitutional schema:

(ITR) ‘T’ is a rigid designator of an entity e iff it is not the case that e might not have been T

“[A]lthough the man (Nixon) might not have been the President, it is not the case that he might not have been Nixon” (1980, p. 49). So ‘Nixon’ seems to satisfy the intuitive test for rigidity. But why doesn’t ‘the President’ satisfy as well? It is true that we often say, in discussing the President, that he might not have been president. The President, we all know, might not have been president; he might have lost. This may seem to prove Kripke’s point. But it does not prove that point.

First, as I have emphasized, Russell distinguishes the issue of whether the President might not have been president from the issue of whether he might not have been the President. Though he might not have been president (if e.g. he had lost), it is not the case, at least in one sense, that he might not have been the President. On Russell’s theory it is simply not the case that if Nixon might not have been president, then (in those circumstances) he would not have been the President. This is a surprising result. But it becomes clearer when understood in terms of (c’) and (d’) above. And to fail to appreciate this result is to misunderstand either the effect of scope or the distinction between the copula and the ‘is’-of-identity.

If it seems implausible to suppose that it is not the case that Nixon might not have been the President, remember that it is false only on one construal. If we give the definite

been much shorter than he actually is) and shorter than Aristotle-as-he-actually-is)]. These are just the interpretations predicted by Russell’s theory. Again, there is a presumption against these interpretations because there is a presumption in favor of giving the names wide scope over the modal operator.
description narrow scope, it is true. And if we misinterpret the definite description as a predicate, again, the sentence can be true—though there is here also a reading on which it is false (see (c')).

For names on Russell’s theory to satisfy (ITR), the theory must have the resources to represent the following sentence (or an analogous sentence—the definite description disguised by ‘Nixon’ could not be simply ‘the President’) as false: Nixon might not have been the President. It does have those resources.

The Residue

I have argued that Kripke’s arguments do not show that Russell’s theory is inadequate to the phenomena. In general, Kripke’s strategy for rejecting that theory is to claim that sentences that the theory would make equivalent are not in fact equivalent. But to deny their equivalence, Kripke relies on one interpretation of sentences that Russell’s theory predicts will be ambiguous. That ambiguity enables the response to Kripke’s claim; a response I have tried to sketch above.

There is, however, a residual philosophical issue. Consider again two sentences Russell’s theory would make equivalent.

(i) Nixon might not have been the President.
(ii) Nixon might not have been Nixon.

Kripke claims that these sentences are not, in fact, equivalent. I have argued that both are ambiguous and that there are available, on Russell’s theory, readings that make them equivalent, as well as readings that make them inequivalent. The residual philosophical issue is that although (i) and (ii) are equivalent on all corresponding disambiguations, it is perhaps most natural to disambiguate them in a way that makes them inequivalent.
According to Kripke’s intuitions, anyway, (i) is true and (ii) is false. If we accommodate those intuitions by giving the description in (i) narrow scope and the second occurrence of ‘Nixon’ in (ii) wide scope, then we have not disambiguated (i) and (ii) in corresponding ways.

But if names are simply disguised or abbreviated definite descriptions, why should there be this difference? There seems still to be a challenge to Russell’s theory. Although there need be no difference between names and definite descriptions as far as their representation in logical form, i.e. sentences that should be logically equivalent can be equivalent according to Russell’s theory, there is still a difference as far as the most natural disambiguation of sentences containing names and definite descriptions.

If two sentences are the same except for containing, respectively, a name and a definite description the name is supposed to abbreviate, then Russell’s theory predicts those sentences will have the same underlying logical form. If those sentences contain a modal operator however, both sentences will be ambiguous. Although their corresponding logical forms will be equivalent, it is most natural to represent the sentences logically as inequivalent. The residual philosophical question: What does this difference show?

Stating the issue clearly leads us to a response. Russell’s theory, I claim, is about the semantic content of certain kinds of terms. The semantic content of a singular term, in general, is the semantic contribution it makes to larger expressions of which it is a part. The semantic content of a sentence as a whole is a proposition—which the sentence says, what we understand when we understand the sentence. Such a theory will be independent of certain special conventions we may have regarding the interpretation of those terms.

The fact is, I believe, that the possibility of interpretative conventions differentially affecting one of two terms is not necessarily a challenge to the semantic
equivalence of those terms. An illustrative example for my purposes could be created with the terms “use” and “utilize.” It would be no criticism of a theory that identified the semantic content of those terms that the utilization (!) of one of the terms is pompous or affected in a way that the use of the other is not. Perhaps even more analogously, ‘and’ and ‘but’ have the same semantics but (!) different conventional implicatures. Similarly, I claim, Russell’s theory that a name has the semantic content of some definite description—better, that a sentence containing a name expresses the same proposition as an otherwise identical sentence in which the name is replaced by a definite description—is unaffected by the phenomenon that names are more naturally interpreted as having wide scope in the context of modal operators. Both scopes are and should be semantically available for both kinds of terms.

In general, we should not expect the kind of equivalence Russell claims to hold between names and definite descriptions to have the consequence that any kind of linguistic rule that applies to one kind of term will apply as well to the other. Names, for example, are supposed to be capitalized.13 In the cases Kripke considers, I would argue, we have conventions that demand or at least prefer syntactic transformations of the relevant sentences that do not correspond to each other. The convention may be as simple as the (defeasible) presumption that the quantifier introduced in the Russellian representation of a sentence containing a name be given wide scope over any modal operator in whose context the name originally occurs. This preliminary formulation is unsophisticated and would need substantial refinement to accommodate cases involving

13 Mr. cummings, Ms. hooks, and the artist formerly known as ‘Prince’ do not abide this supposition.
nested operators, terms occurring anaphorically in the context of an operator, and so on.\textsuperscript{14} But it gives the idea.

\textit{Conclusion}

My conclusion is that we have seen no linguistic phenomena that necessarily conflict with Russell’s theory of descriptions and his thesis that names are simply disguised definite descriptions. Perhaps a name is such that its contribution to the truth conditions of sentences in which it appears does not vary counterfactually. But a definite description taken with wide scope is also such that its contribution to the truth conditions of sentences in which it appears does not vary counterfactually. I thus do not see the variation between the behavior of names and that of definite descriptions that Kripke claims. Kripke confesses that his remarks on this issue are too brief. I agree; but my objections to his remarks do not focus on any incompleteness.

\textsuperscript{14} I am ignoring, for example, significant complications (discussed by Kripke in a seminar at Princeton, Fall 1993) that concern the order of expansion of definite descriptions.
References


