Solving Frege's Puzzle*

Richard G Heck, Jr

Abstract

Our actions are a product of mental states that represent how the world is, how we want it to be, and so forth. Such is central to our conception of ourselves as rational agents. I suppose it is possible that this conception is simply wrong. Perhaps there are, as eliminativists have argued, no such things as beliefs and desires. Or perhaps there are, but these states are merely relations to uninterpreted formulae in some internal computational system. I am going to set such questions aside here and assume that our ordinary conception of our ourselves is not wholly mistaken. The question I want to discuss concerns the role played in this conception by the notion of representation, that is, by representational content. The question is: How must we understand the contents of mental states such as beliefs and desires if those states are to play the causal and explanatory roles envisaged for them?[1]

The question can be illustrated as follows. According to Frege (1984b, pp. 144–5), the reference of a sentence is its truth-value. Frege did not, however, take the content of a sentence to be its truth-values and, relatedly, he did not regard beliefs as relations between thinkers and truth-values. Such a view would widely be regarded as patently absurd. But why?

One answer is that such an account fits poorly with our intuitions about the truth-values of sentences that attribute beliefs. If beliefs were relations to truth-values, it might be said, then "*N* believes that *S*" and "*N* believes that *P*" would have the same truth-value whenever *S* and *P* had the same truth-value. Since each of us surely has at least one belief that is true and one belief that is false, every sentence of the form "*N* believes that *S*" would then be true. That does not accord with intuition. But this is a poor sort of objection. There is no reason to suppose that our

*A somewhat shortened version of this paper appeared in the *Journal of Philosophy* 109 (2012), pp. 132–74.

[1] The puzzle in which we'll be interested can be formulated in various terms, and whether one wants to think of it as concerning psychological explanation, intentional laws, or mental causation is not, I think, critical for the discussion here. I'll usually talk in terms of explanation, as that is how I've tended to think of it myself, but I'll speak in other terms when that seems convenient.

'folk' conception of ourselves as rational agents will survive utterly unchanged as our scientific conception of ourselves improves. Our everyday conception of belief might have to undergo significant reconstruction, and our 'intuitive' judgments might prove in many cases simply to be wrong. One might reply, of course, that rather too many intuitions would then prove false for the ordinary conception just to have been 'reconstructed'. But the more important point is that we must not confuse questions about *the nature of belief* with questions about *the semantics of belief-attribution*. Questions of the former sort lie, ultimately, within the province of cognitive psychology; questions of the latter sort lie within the province of theoretical linguistics. There is no reason to suppose that the semantics of ordinary language belief-attributions must mirror the facts about belief itself: The scientific notion of belief need not correspond precisely with the everyday one.[2] But I need not insist on this point. If there is just one concept of belief, then it is the ordinary concept that figures in the work of cognitive psychologists, and there is no reason to suppose that intuition is a particularly reliable guide to the truth-values of sentences containing the verb 'to believe'. You cannot do cognitive psychology from the armchair.[3] So either way, the conflict with intuition is irrelevant.

Now, to be sure, there are views that imply that questions about the nature of belief cannot come apart from questions about the semantics of belief-attribution. Such a conclusion would follow from any view that took the facts about belief to be, in one way or another, but a reflection of our practices of belief-attribution.[4] But such views will also be left out of account here: The question I want to discuss arises, in the form in which I intend to discuss it, only for broadly realist views of the mind.

What, then, is so obviously objectionable about the psycho-Fregean account of belief as a relation to a truth-value? The answer, I suggest, is that, if beliefs were relations to truth-values, then they could not do the explanatory work required of them. Each of us would surely believe both the truth-values and so believe every-

[2] One might be tempted to object that, since "Fred believes that snow is white" is true if, and only if, Fred believes that snow is white, the facts about belief will impinge upon the semantics of 'to believe' simply because the verb 'to believe' expresses the relation of belief. But this objection is answered just as in the text. The crucial question is whether the verb 'to believe', as it occurs on the right-hand side of the T-sentence, expresses the ordinary concept of belief or the scientific one. If the former, the T-sentence is true but tells us nothing about belief itself; if the latter, it need not be true.

The point is simply that there is just no general reason to suppose that an utterance of "*N* believes that *S*" can be true only if that utterance of *S* has the very same content as some belief of *N*'s. I've discussed this point myself in connection with demonstratives (Heck, 2002), and related points have been emphasized by Herman Cappelen and Ernie Lepore (1997).

[3] I owe my full appreciation of its force to Noam Chomsky (2000).

[4] I have in mind, for example, the views of Donald Davidson (1984b; 1984c). But any interpretavist view—say, that of Daniel Dennett (1971)—will have a similar consequence.

thing there is to believe. But if so, it is more than a little hard to see how my beliefs could explain how I in fact behave as opposed to how anyone else behaves, since we all have the same beliefs; or, perhaps more obviously, how my *not* behaving in certain ways might be explained by my *not* having certain beliefs: Whatever those beliefs might be, I do indeed have them. It is thus the explanatory purposes for which beliefs are needed that rule out the psycho-Fregean view.

My concern here will be with an argument of the same form, but one whose target is not the psycho-Fregean view but the neo-Russellian view that the contents of beliefs are Russellian propositions—or, more precisely, that the contents of beliefs are individuated, so far as the objects the belief is about are concerned, only by those objects themselves.[5] Thus, on this view, the belief that Mark Twain lived in Hartford has the same content as the belief that Samuel Clemens lived in Hartford.[6] And let me emphasize that the view in which I am interested is the view that belief is a *binary* relation between a cognitive agent and a Russellian proposition. This is not a view that is widely held. Many philosophers hold views that lead them to say (or write) things like, "The contents of beliefs are Russellian propositions", when their view is in fact that belief is a *ternary* relation between an agent, a Russellian proposition, and something else, alternately known as a 'guise', a 'mode of presentation', or what have you. It is not at all clear why we should not say, on such views, that the contents of beliefs are Russellian propositions plus 'guises', 'modes of presentation', or what have you. Indeed, it is not clear why a similar maneuver would not salvage the psycho-Fregean account: Perhaps belief is a ternary relation between thinkers, truth-values, and modes of presentation. In any event, we shall return to this issue in section 3.2, and to another form of it in section 4.

To avoid confusion, then, I will refer to the 'pure' Russellian view as the 'naïve' view.

The argument I want to consider, then, goes as follows. Suppose Fred reads in the local paper that Clem Samuels has died. Clem, as it happens, was someone Fred knew from around town, and, while Fred liked Clem well enough, he didn't really know him and so isn't exactly devastated by this news, though it does sadden him. Some time later, however, Fred hears from the local grocer that Sam Clemens has died. Sam was an eccentric neighbor of Fred's, someone of whom he really was very fond. Fred knows already how badly he will miss Sam's strange humor and disarming smile. So Fred is really quite upset about Sam's death.

The fact that Fred was not so upset before the grocer told him about Sam's death but is afterwards seems obviously to be a consequence of his not having

[5] In fact, I'm no fan of structured Russellian propositions, but my reasons for unhappiness are orthogonal to the issues here, so I shall set them aside.

[6] 'Mark Twain' was a nautically-inspired *nomme de plume* used by the American author Samuel Clemens, who did indeed live in Hartford, Connecticut, beginning in 1871.

believed before what he did come to believe then, namely, that Sam Clemens had died. And the naïve view need have no problem with this explanation. Sam Clemens and Clem Samuels were two different people, so, by the naïve theorist's lights, there is no particular reason Fred couldn't have believed that Clem had died without also believing that Sam had died. So, in this respect, the naïve view is an improvement on the psycho-Fregean view, which might have a problem even here. But, of course, there are similar examples that do pose a problem for the naïve theory. We need only suppose that the newspaper from which Fred learned of Clem's death also reported the death of Mark Twain and that Fred, though he knows something of literature and greatly admired Twain's writing, also wouldn't have been personally affected by news of Twain's death. But, familiarly, since Sam Clemens was Mark Twain, the naïve theory implies that the belief that Sam Clemens has died has the same content as the belief that Mark Twain has died so that, even before his conversation with the grocer, Fred *did* believe that Sam Clemens had died. Now the obvious explanation of Fred's change of mood must fail.

Such cases are of course known as *Frege cases*. They pose a problem for the naïve theorist because s'he seems unable to offer any psychological explanation of Fred's change of affect. As Jerry Fodor puts it:[7]

> . . . [*D*]*e re* specifications of propositional attitudes are generally too weak to support explanations of behavior when the latter is intentionally characterized. So, *de re*, Oedipus's desire to marry Jocasta = his desire to marry his mother = his desire to marry the tallest woman in Greece. . . . But it is only the first of these specifications of what Oedipus desired. . . that figures in canonical explanations of the behavior that Oedipus tried/intended to produce. (Fodor, 1982, p. 101)

To elaborate, let t_1 be when Fred read "Mark Twain has died"; and let t_2 be when the grocer told him, "Sam Clemens has died". The change in Fred's mood that occurs at t_2 is plausibly due to some cognitive change Fred has undergone. Of course, it need not be: There are all kinds of reasons Fred's mood might have changed. But the naïve theorist seems committed to the view that this particular change *cannot* have been due to a cognitive change. What makes this so incredible is that it seems to force us to give very different explanations of what would otherwise seem to be very similar occurrences. Suppose, for example, that Fred had never read the paper that morning and so had never read, "Mark Twain has died". Then it would be completely obvious that Fred's change of mood at t_2 was due to a cognitive change: He came to believe that Sam Clemens had died. And, of course, it isn't Fred's reading the paper *per se* that is responsible for the problem the naïve theorist

[7] The particular beliefs Fodor discusses are perhaps not well chosen—the theory of descriptions and all that—but let that pass.

faces: If Fred had read the story about Clem but not the one about Twain, the obvious explanation of Fred's change of mood at t_2 would still have been correct. What is so perplexing, then, is why the fact that Fred did read the story about Twain together with the fact that Mark Twain was Samuel Clemens should not only bar the ordinary explanation of Fred's change of mood but should bar *any* cognitive explanation whatsoever.

The example I've just given concerns, as I have been saying, not a change in Fred's behavior but a change of affect. Other sorts of examples are of course possible, however, including ones that would directly involve a change of behavior, and there are also examples involving other sorts of psychological changes. What all of these examples have in common is that, whatever sort of change might be involved—affective, behavioral, or cognitive—we are ordinarily inclined to explain that change as the result of a change in what the agent believed, and the naïve view seemingly will not permit any such explanation.[8]

The remainder of this paper will be devoted to exploring such arguments in more detail. In the next section, we shall consider Fodor's proposal that we should simply deny that there is any cognitive explanation to be given in such cases—that such cases are, in some sense, pathological. We shall see that this response is every bit as desperate as it seems.[9] In section 2, we'll look at three ways of arguing that the naïve view can offer a cognitive explanation of Fred's change of mood, just not the one we might have expected. We shall see that all of these are insufficiently general: Each falls victim to a Frege case.

The central arguments of the paper are in section 3. I will first extract some lessons from the preceding discussion and use them to argue that the naïve view is indeed untenable: We simply cannot avoid the conclusion that Fred does acquire a new belief when he hears from the grocer that Clemens has died. It does not, however, follow without further argument that this new belief has a different *content* from the belief Fred previously held, that Twain has died, and I will suggest that the naïve view can be resurrected in the form of a denial that these different beliefs have different contents.

This sort of suggestion has been made before, but the problem has always been that it is hard to see how this not-quite-so-naïve view might still provide us with a *cognitive* explanation of Fred's change of mood: Cognitive explanation is supposed to be intentional explanation—explanation in terms of the contents of mental states, as opposed (say) to their neurological properties—and it can easily look as if the fact that Fred's new belief has the same content as his old belief will prevent us

[8] Yet another sort of example would involve changes in what the agent desires: Lex Luthor had previously wanted to destroy Superman, but now he also wants to destroy Clark Kent—not because he believes Clark is Superman but because he believes Clark is passing information to Superman.

[9] Fodor tells me that he has since abandoned it.

from giving any intentional explanation of his behavior. I shall argue that this is an illusion. Much of that argument is contained in section 4, where we shall consider a series of attempts to press the worry that no view that allows a thinker to have different beliefs with the same content can prevent non-intentional features of mental states from playing a significant explanatory role. Perhaps the most important of these objections is the charge that the same moves I make in defense of the naïve view could also be made in defense of the psycho-Fregean view. They can of course be made—but not, I shall argue, effectively.

1 Biting the Bullet

The most radical response to Frege cases is the one Fodor defends in *The Elm and the Expert* (1994):[10] Deny that these kinds of examples really refute the naïve theory. Fodor's idea is that, since psychological laws are not exceptionless, perhaps Frege cases are simply among the exceptions. But the difficutly is that, while there is certainly room in logical space for this view, it is hard to see how it could actually be true. Psychological laws have exceptions, or so one might have supposed, because psychological laws are not basic laws. Psychological states, that is to say, are implemented by states of some more fundamental kind—neurological states, in our case—and hiccups at the neurological level can therefore cause phenomena at the psychological level that are, from the point of view of psychology, anomolous, which is to say that they fall outside the domain of psychological explanation. So the obvious way to construe Fodor's suggestion is this: The very fact that someone has two singular representations of a single object which she does not know to be representations of a single object itself constitutes such a hiccup.

This suggestion is not overwhelmingly plausible. Fodor attempts to reinforce it by arguing that there are mechanisms that will make such occurrences rare. In response, Gabriel Segal (1997) has argued that there are no such mechanisms as Fodor proposes. But I see no reason to suppose that Frege cases are at all unusual. Imagine, for example, that I were sitting outside at a table in Harvard Square. As people walk by, I am visually acquainted with them: I come to have, as Gareth Evans would have put it, demonstrative thoughts about them. But who knows how many of these people are also presented to me in other ways?[11] I know many of

[10] Susan Schneider (2005) has developed another version of this view, arguing that Frege cases may be treated as 'tolerable exceptions' to a *ceteris paribus* law. Her view differs in detail from Fodor's, but I think it is still vulnerable to the objections raised here. That said, her view is in many ways similar in spirit to mine. The central difference is that I do not need the idea that Frege cases are, in any sense, exceptional. See note 33 for some further remarks.

[11] In the course of presenting this paper, several Frege cases were generated: There were, each time, at least a few people in the audience whom I'd not previously met but whose work I'd read, and

my former colleagues by name though not by sight—and while one could debate the frequency of the phenomenon, it hardly seems like the sort of thing that would spring a *ceteris paribus* clause, thus making my behavior towards these people psychologically inexplicable.

The more serious problem, however, is that Fodor's response threatens to deprive us of the ability to give any psychological explanation whatsoever of my coming to know, say, that the person walking down the street is Krister Stendahl.[12] Such discoveries of identity are extremely common. A few weeks ago, for example, I was looking outside the window of my study when I saw a cat who looked very much like my cat, Joe. Joe is an indoor cat. But, as I realized after a minute or two, that cat *was* Joe, who had apparently escaped to the great outdoors. In so recognizing Joe, I was making an identity judgement: *That* creature—the one presented to me visually, in such and such a way—is Joe. One can see that two 'modes of presentation' must be involved here by reflecting on the fact that I did not originally recognize Joe, and the structure of the phenomenon would not have been different had I recognized him immediately. Even if I had, I could intelligibly have wondered whether that creature really was Joe.

The problem, then, is that, on Fodor's view, my arriving at the identity judgement *that creature is Joe* cannot be regarded as a psychological phenomenon admitting of a psychological explanation. But the discovery of such identities is characteristic of object-recognition and so is an almost pervasive feature of human cognition. It would be well beyond desperate to banish such recognition from the realm of psychological explanation.

2 Variations on a Theme by Frege

The sort of argument we are considering here has, of course, been widely discussed in the literature,[13] but it is fair to say, I think, that few have wanted to defend the naïve theory against it. There is a parallel debate in the philosophy of language, though, and there the relevant analogue of the naïve theory—a naïve theory of belief-*attribution*—has been vigorously defended.[14] I've warned against conflating

so forth.

[12] Similar worries are expressed by Jerome Wakefield (2002).

[13] The classic discussions include those of Fodor (1982; 1994), McGinn (1982), Block (1986), and Loar (1988b).

[14] In fact, it isn't always clear whether a particular author is discussing a thesis about belief or, instead, a thesis about belief-attribution. Some people do not distinguish the two theses at all, or they purposely run belief and belief-attribution together because they assume some strong connection between the two or, relatedly, between the contents of mental states and the contents of utterances. For example, Block's desiderata on an acceptable theory of content include both psychological and linguistic elements (Block, 1986, pp. 616ff); McGinn, though he surely must have been aware of the

these two sorts of issues, but one would nonetheless expect that at least some of the moves and counter-moves that are made in the case of belief-attribution could also be made in the case of belief itself. So what I propose to do in this section is to adapt some of the arguments that have been offered in defense of neo-Russellian accounts of belief-attribution and see how they fare when deployed in defense of the naïve theory of belief itself.

In doing so, I am being unfaithful to the intentions of at least some of the authors of these arguments. As David Braun emphasizes in "Russellianism and Explanation" —we'll focus on his discussion in section 2.1—most Russellians about belief-attribution are *not* naïve theorists about belief (Braun, 2001, pp. 256–7).[15] Rather, they regard belief itself as a ternary relation between an agent, a Russellian proposition, and a way of taking (or grasping, or what have you) such a propo-

distinction, essentially ignores it (McGinn, 1982); and Loar, though in a sense his entire discussion is concerned with this distinction between belief and attribution, obscures it through his emphasis on "*commonsense* psychological explanation" (Loar, 1988b, p. 99, his emphasis). I am somewhat guilty myself, and David Braun seems to have been misled by my sloppiness. He claims that I "explicitly use [Frege cases] to argue against Russellianism" (Braun, 2001, p. 287, fn. 47), that is, against a Russellian view about attitude *ascriptions*. But my concern in the passage Braun cites is not with attitude ascriptions but with attitudes (Heck, 1995, pp. 79–80). I do not there "argue[] that, if Russellianism were true, then attitude ascriptions could not be used to explain why certain agents... do not behave in certain ways" (emphasis removed) but, rather, that, if the naïve theory were true, agents' *having certain attitudes* would not explain their failing to behave in certain ways. That said, although I emphasize that the issue "does not only concern intuitions about belief reports", I then confuse matters by saying, much as Loar does, that what is at stake is "the status of everyday explanations of behavior" (Heck, 1995, p. 80, fn. 4), by which I seem to mean the sorts of explanations that might be given by ordinary speakers. If so, however, it might look as if what mattered was the truth of the attributions made in the course of such explanations, whence the issue might well seem to be one about semantics. That is not, in fact, what I had in mind: The paper as a whole is concerned with the question how the contents of attitudes are related to the contents of sentences, and it purports to uncover a tension in any view that marries a broadly Fregean account of the attitudes to a broadly Russellian account of proper names. The point of the remarks Braun cites is pretty clearly to explain why I took it to be relatively uncontroversial that Frege was right about the contents of belief even though it is exceedingly controversial whether he was right about the contents of sentences. The semantics of attitude ascriptions are hardly mentioned.

All of this is a good deal clearer in "Do Demonstratives Have Senses?" (Heck, 2002), to which Braun would not then have had access.

[15] There is a certain irony in this, since Russell himself most emphatically was a naïve theorist. This point is sometimes obscured by the fact that Russell thought us capable of so few singular thoughts. But where singular thought is possible, Russell insisted, the naïve theory is true. And, indeed, the existence of Frege cases was one of the reasons Russell limited the extent of singular thought as he did.

Since I am going to be fairly critical of Braun, let me say explicitly that his papers "Russellianism and Explanation" (Braun, 2001) and "Russellianism and Psychological Generalizations" (Braun, 2000) are what inspired me to think more deeply than I previously had about how a naïve theorist might respond to Frege cases—even though Braun's papers are not really concerned with that issue.

sition. Now, I certainly do not deny that this combination of views is available: Since the metaphysics of belief is one thing and the semantics of belief-attribution is another, it certainly is available. But it is too little noticed what a difficult position ordinary speakers would be in were such a combination of views correct. It might well be, for example, that Fred's becoming upset when he was told "Sam Clemens has died" was due to some cognitive change he had undergone: He came to have a new belief; that is, he came to stand in the belief-relation to a proposition and a way of taking that proposition in which he had not stood before. Unfortunately, the alleged facts about the semantics of attitude verbs (or, perhaps better, about the semantics of complement clauses) would then prevent ordinary speakers from reporting that such a change had occurred in the terms in which, as a matter of obvious fact, they actually try to report such things: It would simply be false to say "Fred then came to believe that Sam Clemens had died, whereas he had not so believed beforehand". And though it may well be that sentences that explicitly mention ways of taking propositions could be used in giving such explanations (Braun, 2001, pp. 278–9), using such sentences would require knowing which ones they were, and even we philosophers do not actually know how to mention ways of taking propositions. Such sentences are well beyond the ken of ordinary speakers, in any event. So, even if it is not inconsistent, the combination of a Russellian view of attitude ascription and a ternary view about belief seems to me to border on incoherent.[16]

More importantly, someone who holds this combination of views presumably thinks that there are reasons to do so, that is, that there are reasons to endorse a ternary metaphysics of belief. Braun does not say what those reasons are, but the sort of argument most frequently given (and cited) is just the broadly Fregean argument we are presently considering.[17] That fact does not make Braun's position inconsistent—facts about belief are, once again, different from facts about belief-attribution—but it does make the position dialectically uncomfortable. If, on the one hand, our adaptation of Braun's response to Frege's puzzle about belief-attribution were to succeed, we would then be left without any good reason to affirm the account of belief that Braun endorses; but if, on the other hand, our adaptation failed, then the question would naturally arise whether Braun's original response did not suffer from analogues of the problems we would then have discov-

[16] The obvious reply is that, although ordinary speakers cannot *say* that such a change has occurred by uttering "Fred then came to believe…", they can *communicate* that such a change has occurred by uttering this sentence. But can "Fred became upset at t_2 because he then came to believe that Clemens had died" be true if what follows "because" is false? Isn't causal "because" factive? I'm sure there are responses, but I still find the view incredible.

[17] One widely cited discussion, for example, is Nathan Salmon's, in *Frege's Puzzle* (Salmon, 1986, ch. 8).

ered with the adapation.[18] And most other philosophers who endorse a Russellian treatment of the attitudes are in the same pickle.[19]

But however that may be, it is worth seeing whether one of the extant defenses of neo-Russellian accounts of belief-attribution can be adapted to defend the naïve theory of belief itself. The challenge, recall, is to identify the cognitive change in Fred that is responsible for his change of mood when he is told "Sam Clemens has died". The naïve theorist cannot say that Fred then comes to have a belief with the content *Sam Clemens has died* when he did not previously have a belief with that content. But perhaps something else has changed.

2.1 The Braun Variation

David Braun's suggestion (adapted to the case of belief) is that we should explain Fred's change of mood in terms of his having come to believe at t_2 that *his eccentric neighbor* has died, whereas he did not believe this at t_1 (Braun, 2001, p. 277).[20] The naïve theory is clearly consistent with this claim.[21]

Before we discuss this suggestion further, I want to note an odd consequence of it: If this response is adequate, then the crucial cognitive difference in Frege cases must always lie in agents' *descriptive* beliefs, since the naïve theorist can never allow, in such a case, that there is a difference in the agents' *singular* beliefs. Frege cases are (real or hypothetical) situations in which we are inclined to explain someone's behavior, or other mental states, by saying that, although s'he does believe that $F(a)$, s'he does not believe that $F(b)$, even though $a = b$, which fact of course prohibits the naïve theorist from endorsing this explanation. But—or so it seems to me—if the agent's singular beliefs do not explain her behavior in Frege cases, then it is unclear how they can explain her behavior in non-Frege cases, either.[22] Are we to say, for example, that it *would* have been appropriate to explain Fred's becoming so upset at t_2 in terms of his then coming to believe that Sam Clemens had died if he had not previously read "Mark Twain has died" but that, since he did previously read "Mark Twain has died", it is inappropriate to do so? I do not say that such a view could not be ably defended. I do say that such a view is not

[18] The adaptation will indeed fail—see section 2.1—and, so far as I can see, there is nothing to prevent us from raising a parallel objection to Braun's actual discussion of attribution.

[19] To be sure, there is at least one way of responding to Fregean concerns about belief-attribution that does not neatly transfer to the case of belief, namely, pragmatic responses. But these have their own problems, as Braun (1998) shows. See note 16 for a bit more on this.

[20] A similar view, directed at the case of belief, is found in Michael Thau's *Consciousness and Cognition* (2002).

[21] Assuming, that is, that "his eccentric neighbor" is not a referring phrase but a definite description to be treated *a la* Russell.

[22] This kind of worry figures extensively in Segal's writings on this issue (2000a; 2000b).

likely to be true. If so, however, then the naïve theory, defended in the way we are presently considering, feels a good deal more like Russell's actual epistemology than I would have supposed its proponents intended.[23]

But even if we waive that point, it is easy to see that the response we have adapted from Braun will not deal with all Frege cases. We need only ask *why* Fred comes to believe at t_2 that his eccentric neighbor has died when he did not so believe at t_1. It is not as if it was because of a bump on the head. Fred's being told "Sam Clemens has died" is surely part of the story: Had the grocer instead said, "We have apples on sale today", Fred would not have come to believe that his eccentric neighbor had died. Similarly, if Fred had not believed that Sam Clemens was his eccentric neighbor, he would not have come to believe that his eccentric neighbor had died, either. So both Fred's pre-existing belief that Sam Clemens was his eccentric neighbor and his being told "Sam Clemens has died" played a role in Fred's coming to believe that his eccentric neighbor had died. How do these fit together? The obvious answer is that, when Fred was told "Sam Clemens has died", he came to believe that Sam Clemens had died, and then, knowing that Sam Clemens was his eccentric neighbor, he inferred that his eccentric neighbor had died. But this answer is obviously not available to the naïve theorist, since, according to h'er, Fred already believed at t_1 that Sam Clemens had died, and he had believed all along that Sam Clemens was his eccentric neighbor.

As I mentioned earlier, many presentations of Frege cases focus on the explanation of an agent's behavior: Why doesn't Sally, obsessed as she is with Rosalind Smith, rush to get an autograph when she sees the sign announcing "Book Signing with Joyce Carol Oates"?[24] The case with which we began concerns a change of affect: Why does Fred only become so upset when he is told "Sam Clemens has died"? The case we've just discussed concerns a *cognitive* change, but it is otherwise similar: Why does Fred only come to believe that his eccentric neighbor has died when he is told "Sam Clemens has died"? So the objection I have raised here is that the response adapted from Braun is insufficiently general: There are Frege cases to which it simply does not apply. Worse, it generates a Frege case to which it does not apply.

[23] Later (see page 42), I shall discuss an objection to my own view that has a somewhat similar feel. My response to that objection depends upon the assumption that Fred *does* acquire a new singular belief when he is told "Clemens has died", and the response we are adapting from Braun denies this.

[24] This sort of example is due to Lewis (1986, pp. 58–9) and is discussed by Braun (2001, pp. 273ff) under the title "The Contrastive Explanation Objection".

2.2 The Meta-Linguistic Variation

Another option is to go meta-linguistic: When Fred is told "Sam Clemens has died", he comes to believe that the person named 'Sam Clemens' has died.

The immediate problem with this proposal, if we intend to adapt it to the present context, is that it is far from obvious why Fred's believing that the person named 'Sam Clemens' has died should cause him to become so upset when his believing that Sam Clemens had died did not. A similar question might have been raised about Braun's proposal, but a reasonable answer would also have been available: The story was that Fred would miss Sam's odd humor and disarming smile, and one might suggest that it is *his neighbor's* humor and *his neighbor's* smile that he would miss—though this does, once again, raise the worry that the view being defended is becoming more like that of the historical Russell than we might have wanted it to be. No such response seems available in the present case, however.[25] It isn't as if Fred is particularly attached to the *name* 'Samuel Clemens'.

It might be suggested, however, that we would get something more promising if we were to weld the meta-linguistic response to Braun's. What the meta-linguistic proposal offers us, the thought might be, is an answer to the question why Fred comes to believe at t_2 that his eccentric neighbor has died: He infers that his neighbor has died from his pre-existing belief that his neighbor is named 'Sam Clemens' together with his newly acquired belief that the person named 'Sam Clemens' has died.

I suggest one last time that we are on our way to dispensing with any significant role for singular thought in human cognition. But this maneouver doesn't save Braun's proposal, anyway. The objection offered in the last section was that, if we ask why Fred comes at t_2 to believe that his neighbor has died, we get a Frege case Braun's proposal cannot handle. But we might just as well have asked a slightly different question. Fred already believed at t_1 that Sam Clemens was his eccentric neighbor. So according to the naïve theory, he also believed at t_1 that Mark Twain was his eccentric neighbor. But it is uncontroversial that Fred believed at t_1 that Twain had died. So why didn't he believe already at t_1 that his neighbor had died?

The same sort of objection can be lodged against the meta-linguistic proposal. The idea, recall, was that we should explain Fred's change of mood at t_2 in terms of his then coming to believe that the person named 'Sam Clemens' had died. But we can surely suppose that Fred knew already at t_1 that Sam Clemens was named 'Sam Clemens', and the naïve theorist is therefore committed to holding that he knew as well that Mark Twain was named 'Sam Clemens'. Fred then came to believe, at t_1, that Mark Twain had died. So why didn't Fred believe already at t_1 that the

[25] Similar points have been made by William Taschek (1992, pp. 782–3).

person named 'Sam Clemens' had died? This is just another Frege case, and the meta-linguistic proposal cannot resolve it.

Two points about this argument. First, one might worry that it depends upon some sort of closure principle, say, the assumption that, if Fred believes at t both that p and that q, and if p and q together entail r, then Fred must also believe at t that r; but it is surely not to be expected that people should believe all the consequences of their beliefs, even the simple logical consequences.[26] But I am not assuming any such principle. It is not just that Fred *does not* infer at t_1 that the person named 'Sam Clemens' has died, and it is not as if the reason he does not do so is that he is psychologically incapable of considering all the premises of the inference together. We can imagine that Fred, even while he is fully and consciously aware that Mark Twain has died, is quietly wondering to himself whether the person named 'Sam Clemens'—who he is also fully and consciously aware is just Sam Clemens—might also have died. Fred is in *no position* to infer that the person named 'Sam Clemens' has died and, if he did so infer, his newly formed belief would be unjustified, even irrational. What we lack, then, is any plausible account of why Fred should have been unable to make this inference, rationally speaking.

Second, the argument does not depend upon the existence of any sort of connection between explanation and generalization. It might seem otherwise.[27] It might, in particular, seem as if the argument depended upon the assumption that:

> Given Fred's pre-existing belief that Sam Clemens is his eccentric neighbor, if what explains Fred's coming to believe that his eccentric neighbor has died is his coming to believe that Sam Clemens has died, then, in general, if Fred comes to believe that Sam Clemens has died, he will (at least be in a position to) come to believe that his eccentric neighbor has died.

The naïve theorist might then insist that there is no such connection between explanation and psychological generalization. But the argument I am developing simply does not depend upon any such assumption. The crucial assumption in my

[26] The contrary view has been held: That belief is closed under logical consequence follows from the claim that the contents of beliefs are sets of metaphysically possible worlds. So see, for example, Robert Stalnaker's *Inquiry* (1984) and his two papers on logical omniscience (Stalnaker, 1999b,c) for discussion. This view of course gives rise to problems not unlike those we are discussing here, and it is a natural question whether the resources I deploy would not also be available to Stalnaker. I shall discuss this point below (see page 46), where I argue that they are not.

[27] The suggestion that it does is adapted from Braun (2001, p. 259). I am not sure whether Braun himself would want to pursue this response in the present case, however, because the argument I am developing here seems to be a form of what he calls the 'Contrastive Explanation Objection', against which he does not deploy this kind of response, rather than a form of the 'Ordinary Explanation Objection', where he does deploy it.

argument is that, when Fred comes to believe at t_2 that his eccentric neighbor has died, his doing so is to be explained by his having undergone some other cognitive change, and the question we are pursuing here is what other cognitive change the naïve theorist can identify. One can deny this 'crucial assumption'—deny, that is, that Fred's change of mind is due to some other cognitive change—but doing so amounts to endorsing the response of Fodor's we discussed in the section 1, and I have already argued that Fodor's response will not do.

2.3 The Soames Variation

As said, it is not as if Fred suffers from some psychiatric condition that isolates his belief that Mark Twain has died from his belief that Sam Clemens is his eccentric neighbor. But it might be suggested that Fred suffers from a different sort of compartmentalization. Scott Soames (1987, pp. 221ff) makes precisely this kind of suggestion in connection with belief-attribution: We should, Soames says, distinguish between an attribution of the two beliefs that a is F and that a is G and an attribution of the single belief that a is both F and G. Adapting this suggestion to our present concerns, then, a naïve theorist might suggest that, while Fred does indeed believe at t_1 both that Sam Clemens has died and that Sam Clemens is his neighbor, he does *not* believe at t_1 that Sam Clemens has died and is his neighbor. The latter belief is one Fred only acquires at t_2, and it is this belief that explains his being so upset.

This idea could also be used to explain why Fred only comes to believe at t_2 that his neighbor has died. The thought would be that Fred can infer that his neighbor has died only from a single belief of the form: N has died and is Fred's neighbor, not from distinct beliefs of the forms: N has died; and: N is Fred's neighbor. And, in principle, the view might also hope to avoid the other problems that beset the earlier proposals. One might claim, for example, that it is Fred's single belief that Sam Clemens has died and is his neighbor that explains his being so upset—though I, for one, would want to hear more about this. But it doesn't really matter, since this proposal too falls to a Frege case.

The obvious question to ask is *why* Fred comes at t_2 to believe that Sam Clemens has died and is his eccentric neighbor. Simply to say that Fred's beliefs had previously been compartmentalized in some as yet unexplained way is no answer. What led to Fred's having the conjunctive belief was the grocer's telling him "Sam Clemens has died". The obvious way to explain why he then comes to believe that Sam Clemens has died and is his neighbor when he had not previously so believed is in terms of his then coming to believe that Sam Clemens has died: He previously believed that Sam Clemens was his neighbor; he came to believe that Sam Clemens had died; and he made the obvious inference. So this is a Frege case: There is a

change— a cognitive change, in this case—that we are inclined to explain in terms of Fred's coming to believe that $F(b)$, but he had antecedently believed that $F(a)$, so, since $a = b$, the naïve theorist is prohibited from endorsing this explanation.

A different way to press the same point is as follows. Consider Fred at t_1 and suppose with the Soames-inspired naïve theorist that Fred believes both that Sam Clemens has died and that Sam Clemens is his neighbor but does not believe that Sam Clemens has died and is his neighbor. Question: What needs to happen for Fred to get himself into a position to acquire this latter belief? Of course, Fred might learn directly that Sam Clemens has died and is his neighbor. Someone might tell him precisely that. But surely that is not the only way Fred might arrive at this belief, and in the present case it would seem to be Fred's being told "Sam Clemens has died" that did the trick. But how?[28] There are plenty of cases in which someone who knows that a is F acquires the new information that a is G and all but simultaneously acquires the new information that a is both F and G. And then there are the cases in which the latter information is not only not simultaneously acquired but seems to be wholly unavailable. What is the difference between these two sorts of cases? It is not, again, as if the difference is due to a serotonin imbalance. The difference is *cognitive*, and the problem for the naïve theorist is that there seems to be no cognitive difference between such situations that she can identify.

There is really a more basic question to be asked, namely, how we are supposed to understand the inference from the two beliefs that N is F and that N is G to the single belief that N is both F and G.[29] Under what circumstances will such an inference will be available to a thinker? The sort of explanation we have borrowed from Soames obviously will not help us answer that question. But if we knew when such inferences were available, then we would not need the Soames-inspired

[28] In principle, of course, one could mix and match the three responses we are discussing: One might say that some Frege cases are to be resolved Braun's way, some *via* semantic ascent, and some Soames's way. But mixing and matching doesn't seem to help here. For example, one could try saying that it is because Fred came to believe at t_2 that the person named 'Sam Clemens' had died that he then came to believe that Sam Clemens had died and was his neighbor. But it will then be asked why Fred did not previously know that the person named 'Sam Clemens' had died. Is the answer that, though Fred knew both that Clemens was named 'Clemens' and that Clemens had died, he did not know that Clemens was named 'Clemens' and had died? But then, what would he have had to learn to acquire this last bit of knowledge? Once again, his being told "Clemens has died" is what seems to have done the trick, but it's entirely unclear how.

[29] Such inferences have to be possible: We cannot simply make do with single-premise inferences. But the problem isn't limited to multi-premise inferences. We need to understand when, say, the inference from N is F to N is either F or G is valid, that is, when the premise and conclusion are 'appropriately related'. Fred believes that Twain is an author. So why can't he conclude that Clemens is either an author or a spy and then, since he is firmly convinced that Clemens is not a spy, infer that Clemens is an author? Indeed, the problem arises just with the inference from 'Twain is an author' to 'Clemens is an author'. But now I'm getting ahead of myself.

machinery. The work it does would already be done by our account of when the belief that N is F and the belief that N is G were so related that this sort of inference was possible. For example, we would not need to say that Fred can infer that his eccentric neighbor has died only from the single belief that Sam Clemens has died and is his neighbor but, rather, that he can make this inference when the two beliefs that Sam Clemens has died and that Sam Clemens is his neighbor are appropriately related and cannot make it otherwise. But nothing we have yet seen gives us any indication what it might be for these beliefs to be 'appropriately related'.

2.4 Frege on 'Proper Knowledge'

In the next section, we shall consider a further reply on behalf of the naïve theorist, one that is not vulnerable to variations on what should now be a familiar Fregean theme. Before we do so, however, it is worth pausing to note that we are now in a position to offer a reasonable interpretation of one of Frege's more enigmatic remarks on these issues.

In "On Sense and Reference", Frege famously argues against an account of identity-statements that regards them as expressing meta-linguistic claims to the effect that the names occurring in the statement have the same reference.[30] Frege's argument is this:

> . . . [T]his relation would hold between the names or signs only in so far as they named or designated something. It would be mediated by the connection of each of the two signs with the same designated thing. But this is arbitrary. Nobody can be forbidden to use any arbitrary producible event or object as a sign for something. In that case, the sentence $a = b$ would no longer refer to the subject matter, but only to its mode of designation; we would express no proper knowledge by its means. (Frege, 1984c, p. 157, original p. 26)

What does Frege mean by 'proper knowledge'?

David Kaplan has suggested, plausibly enough, that what Frege means is knowledge about the world, in particular, non-linguistic knowledge—except in those special cases where linguistic items are clearly at issue [[FIXME: REF]]. Frege is insisting, for example, that the sentence "Hesperus is Phosphorous" expresses astronomical knowledge, not just knowledge about names. But someone might reasonably reply, it seems to me, that "'Hesperus' and 'Phosphorous' co-refer" *does* express astronomical knowledge since, after all, both 'Hesperus' and 'Phosphorous' refer to heavenly bodies.

[30] This interpretation has been challenged by Michael Thau and Ben Caplan (2001). I am not convinced (Heck, 2003).

I do not know how Frege would have answered this objection, but the following reply falls out of our discussion. Consider Fred again. Fred is largely ignorant of astronomy, but he does at least know that 'Hesperus' refers to Hesperus and that 'Phosphorous' refers to Phosphorous.[31] According to the naïve theory, then, Fred also knows that 'Hesperus' refers to Phosphorous. Now, to be sure, it does not follow that Fred knows (or even believes) that both 'Hesperus' and 'Phosphorous' refer to Phosphorous, so it does not follow that Fred knows (or even believes) that 'Hesperus' and 'Phosphorous' co-refer. But it's not as if Fred doesn't have these beliefs because he hasn't gotten around to making certain obvious inferences. It is rather that Fred seems to be in no position to make the relevant inferences. The question worth asking is thus what Fred would need to learn before he could make the inference in question and so arrive at these beliefs. Or, to put the point differently, and in a way somewhat more reminiscent of Frege's own language: How can we represent the discovery that Hesperus is Phosphorous as a *scientific achievement*?

I don't myself know how Pythagoras discovered that Hesperus and Phosphorous are one and the same. But here is one possibility. Perhaps Pythagoras was able to plot the positions of both of them reasonably accurately and then realized that the position his calculations predicted Hesperus would occupy at some time t was the same as the position predicted for Phosphorous at t. But the naïve theorist cannot endorse this explanation. For suppose Pythagoras had completed his calculations concerning Hesperus and so knew that Hesperus would be at location l at time t. According to the naïve theorist, Pythagoras therefore knew as well that Phosphorous would be at location l at time t. Why, then, did he need to do another set of calculations? One might suggest applying Soames's suggestion to this case: Perhaps we should say that what Pythagoras needed to know was that both Hesperus and Phosphorous would be at l at t but that all he would have known, had he not done additional calculations, was that Heperus would be at l at t and that Phosphorous would be at l at t. But how does doing the additional set of calculations help? One would ordinarily have supposed that it was only after doing the second set of calculations that Pythagoras knew that Phosphorous would be at l at t. But, of course, the naïve theorist cannot agree.[32]

[31] One might be tempted to deny Fred even this knowledge, claiming that he does not really know that 'Hesperus' refers to Hesperus but only that 'Hesperus' refers to the object to which 'Hesperus' refers. But I can see no way of defending this view except by denying Fred any knowledge of objects that is not mediated by the expressions he uses to refer to them, so that Fred doesn't really know that Hesperus is far away, either, only that the object to which 'Hesperus' refers is far away. But that, once again, is just a form of Russell's epistemology, one that makes the descriptions through which we know the world meta-linguistic.

[32] And even waiving that point, surely Pythagoras did know after doing the first set of calculations that both Hesperus and Hesperus would be at l at t. But then, according to the naïve theorist, he

So the problem at which Frege was gesturing, I suggest, is that there seems to be no way for the naïve theorist to explain, in broadly cognitive terms, how Pythagoras might have arrived at his discovery nor why, when he did, it might have counted as rational, justified, or knowledgable.

3 Solving Frege's Puzzle

In the previous section, we considered a variety of responses on behalf of the naïve theorist to the challenge posed by Frege cases. All of these were frustrated by essentially the same problem, namely, that we have no way to insulate Fred's beliefs about Sam Clemens from his beliefs about Mark Twain—for example, to insulate his belief that Clemens is his eccentric neighbor from his belief that Twain has died. As noted in section 2.3, it is tempting to say that these beliefs are compartmentalized somehow, but they are not compartmentalized in any familiar sense: Fred can perfectly well have his belief that Mark Twain has died firmly in mind and simultaneously be wondering whether Sam Clemens has died. Or again: Fred can simultaneously be thinking that Mark Twain has died and that, if Sam Clemens has died, then p, without having any inclination whatsoever to conclude that p—and, more importantly, while correctly regarding any belief at which he might so arrive as irrational and unjustified.

The question the naïve theorist is unable to answer is what cognitive change Fred undergoes when he is told "Sam Clemens has died", that being the cognitive change that is responsible for his becoming upset. The responses examined earlier all deny what one is otherwise inclined to say, namely, that when Fred is told "Sam Clemens has died", he comes to have a new belief, a belief I shall henceforth describe—for want of better language—as the belief that Sam Clemens has died. I suggest, therefore, that the lesson we must draw from our discussion so far is that we simply cannot avoid this conclusion: Fred's belief that Clemens has died is a *new* belief, in which case it is also a *different* belief from his previously held belief that Mark Twain has died. If so, however, belief cannot simply be a relation between a thinker and a Russellian proposition. The naïve theory of belief is false.

What follows, then, will not constitute a defense of the naïve theory of belief. What I am going to argue, rather, is that there is nothing to the *contents* of beliefs beyond a Russellian proposition. And I will argue further that, whatever more there is to the intrinsic nature of a particular belief beyond its relating a thinker to a Russellian proposition, this additional material does not play any role in psychological explanation. The view I shall defend, that is to say, has two parts:.

already knew that both Hesperus and Phosphorous would be at l at t, and he was wasting his time.

1. What distinguishes the belief that Clemens has died from the belief that Twain has died is nothing intensional. In particular, these beliefs have the same content.

2. If we are to be able to explain Fred's behavior in cognitive terms, there must be some difference between these beliefs that plays a role in psychological explanation. But no *intrinsic* difference between these beliefs plays that role. The explanatorily relevant difference is an extrinsic, relational one. It concerns how these beliefs are related to other of Fred's beliefs.

This, I suggest, is as close to the naïve theory as it is possible to get once it has been conceded that a single thinker can have distinct beliefs with the same Russellian content. So I shall continue to speak of the 'naïve theory', but the naïve theorist shall henceforth be defending the position just described rather than the original form of the naïve theory.

Let us start with the question whether the position I intend to defend is even available. For a long time, I thought there was a very short argument to the conclusion that it is not. Since Fred's belief that Mark Twain has died and his belief that Sam Clemens has died are different beliefs, they must have different contents, whence the contents must be individuated more finely than Russellian propositions are. This move is an extremely natural one, often tacitly made. What lies behind it, it seems to me, is the thought that beliefs *just are* relations between thinkers and contents. And what lies behind that thought, I suggest, is the view that psychological explanation (in the sense in which we are concerned with it) is intentional explanation, that is, explanation in terms of the contents of psychological states as opposed, say, to the neurological properties of such states.[33] Consider, for example, the following remarks by Brian Loar:

> By *psychological content* I shall mean whatever individuates beliefs and other propositional attitudes in commonsense psychological explanation, so that they explanatorily interact with each other and with other factors such as perception in familiar ways. (Loar, 1988b, p. 99; see also pp. 103, 105, and 197, fn. 8)
>
> [Psychological content] is that content-like aspect of thoughts, of how we conceive things, by reference to which we consider whether com-

[33] Schneider discusses at some length why one might suppose that differences of the sort at issue here must be due to differences of content (Schneider, 2005, §4). She does not, however, discuss the worry that what she calls 'computational explanation', since it isn't intentional explanation, isn't psychological explanation and so fails to make Fred's behavior intelligible as that of a rational agent. I think this is the really serious worry—I think it's what's bothering Fodor, too—and nothing Schneider says even begins to address it.

> binations of them are rational, whether they motivate a given belief or action, and so on. (Loar, 1988a, p. 127, emphasis removed)

The thought, I take it, is that it is all but analytic of the notion of psychological content that the causal and explanatory properties of a belief (*qua* belief) are determined by its content. If so, however, then it would seem to be impossible for Fred's belief that Mark Twain has died and his belief that Sam Clemens has died to be different beliefs with the same content. These beliefs are differently implicated in psychological explanations of Fred's behavior.[34] If they have the same content, however, then we cannot answer the question "whether they motivate a given belief or action" (Loar, 1988a, p. 127) simply in terms of their content. How beliefs and other psychological states "explanatorily interact[ed]" (Loar, 1988b, p. 99) would therefore depend upon more than just their content, and psychological explanation could not proceed simply with reference to the intentional features of psychological states but would have to advert to at least some of their non-intentional properties. Or, as Fodor puts a closely related point: It would seem to follow that psychological laws cannot subsume psychological states simply in virtue of their having the contents they do but must make reference to some other features of those states. And that would be tantamount to denying that psychological laws are intentional in the relevant sense (Fodor 1994, ch. 1, esp. pp. 22–3; compare Fodor 1982, pp. 100–2).

I am now inclined to resist this line of argument, however, or, rather, to resist the conclusion I was previously inclined to draw from it. We do, indeed, have to allow that psychological explanation needs to make reference to features of mental states beyond their content, if content is construed as the naïve theorist would have us construe it. But I will argue that this concession, in the form in which it has to be made, is far less threatening than it has seemed it must be.

3.1 Sense and Psychological Explanation

Before I continue, though, I want to suggest that the problem here may be one with which Fregeans too have to struggle: A parallel problem would arise for Fregeans if it were possible for a thinker to have two different beliefs with the same Fregean thought as their contents.[35] Now, to be sure, Fregeans have generally regarded it

[34] To deny this—that the beliefs are differently implicated in *psychological* explanations—would be to endorse the bullet-biting response considered and rejected in section 1.

[35] Other examples familiar from the literature can be seen as suggesting a similar conclusion. So-called Mates cases (Mates, 1952) are often so understood. And Paderewski cases seem originally to have been intended to cast doubt upon Fregean accounts, since they suggest that someone could understand two different expressions that had the same sense and so both believe and fail to believe a single Fregean thought (Kripke, 1976). Both arguments assume, however, that the notion of sense is

as a constraint on an acceptable theory of sense that this should not be possible, but it is one thing to have a constraint and another to have a theory that actually meets that constraint. And so far as I can see, the only theory that has any chance of meeting it is one we do not want.

Consider the view defended by Christopher Peacocke in *A Study of Concepts* (1992). On this view, the content of a concept is fixed by the possession conditions associated with that concept, and typical possession conditions will make reference only to a small subset of the inferences in which a given concept might be involved. To possess the concept of disjunction, for example, one need only be inclined to regard inferences that have the form of the usual introduction- and elimination-rules for disjunction[36] as 'primitively compelling' and to do so because the inferences have that very form. Now suppose that there were a language that contained two symbols, $\vee$ and $\curlyvee$, both of which satisfied the truth-table for disjunction. Then the mere existence of these different symbols seems to imply that someone might believe what s'he would express using $A \vee B$ but not what she would express using $A \curlyvee B$.[37] And so far as I can see, that is quite compatible with the assumption that the concepts expressed by the two symbols share a possession condition, namely, the one just mentioned. If so, however, then Peacocke seems to be committed to the claim that these two symbols would express the very same concept, and speakers of this language could have different beliefs—those expressed by $A \vee B$ and $A \curlyvee B$, respectively—that had the very same content.

It is tempting to appeal here to the following technical fact: Given any two symbols $\circ$ and $\bullet$ both of which satisfy the usual introduction- and elimination-rules for disjunction, we can prove, in general, that they are equivalent, in the sense that, given any two sentences A and B, we can prove $A \circ B \equiv A \bullet B$.[38] The idea would be that anyone who satisfied the possession conditions for both $\vee$ and $\curlyvee$ would have to realize that they were equivalent and so would never actually believe that $A \vee B$ but not that $A \curlyvee B$—except, perhaps, in the uninteresting case in which she simply

closely tied to public language (see Dummett, 1978), and that can be denied—as, in fact, it routinely is. For a bit more on this, see note 42.

[36] These are the rules $A \vdash A \vee B$ and $B \vdash A \vee B$ together with the disjunctive syllogism: From $\Gamma, A \vdash C$, $\Delta, B \vdash C$, and $\Theta \vdash A \vee B$, conclude $\Gamma, \Delta, \Theta \vdash C$. It does not, of course, matter whether these are the right rules. The sort of worry expressed in the text will arise no matter what they might be.

[37] I.e., you can get Paderewski cases here, too.

[38] One might, indeed, require that a set of inference rules satisfy this condition if they are to determine a concept. As is now well-known, some such restrictions are required: One does not want to suppose that an arbitrary set of rules determines a concept, since some such sets are inconsistent.

The technical situation here is more subtle than is usually recognized. There are ways of combining constants from different logics that prevent the sort of 'collapse' mentioned in the text. One now familiar method is by 'fibring' (Gabbay, 1998). Josh Schechter has developed another, in work as yet unpublished. The point does not affect the present discussion, but it definitely affects other uses to which collapse results have been put.

hadn't bothered to draw the relevant inference.[39] But there are several problems with this suggestion. First, the proof that $A \vee B \equiv A \curlyvee B$, while not exactly difficult, is not utterly trivial, either. (Exercise!) That $A \vee B \equiv A \curlyvee B$ therefore seems like something one could *discover*—and the point of using that word is to emphasize that the Fregean will then have the same sort of obligation in this case that the naïve theorist has in Frege cases, namely, an obligation to offer a *psychological* account of this discovery. Second, it is important to understand that the 'proof' of which I've been speaking is meta-theoretic. It is a proof *schema* that encapsulates a method for constructing a proof of $A \vee B \equiv A \curlyvee B$, given any sentences A and B. The meta-theoretic proof therefore licenses us to conclude that, for all sentences A and B, $A \vee B$ is equivalent to $A \curlyvee B$. But the meta-theoretic proof, and so the meta-theoretic conclusion, need not be available to the ordinary thinker who possesses the concepts expressed by $\vee$ and $\curlyvee$: No capacity for meta-theoretic reasoning (or, if you prefer, schematic reasoning) is required for possession of such concepts. The ordinary thinker, then, need not even be able to entertain the thought that $A \vee B$ is always equivalent to $A \curlyvee B$. The ordinary thinker is therefore not in the same position we theorists are: Knowing the meta-theoretic result, we might reasonably regard $\vee$ and $\curlyvee$ as mere synonyms; but the ordinary thinker need not know the meta-theoretic result nor, as I said, even be able to entertain it.

One common reaction to this kind of problem is to say that such a thinker would have only an incomplete grasp of the concepts expressed by $\vee$ and $\curlyvee$. Perhaps that is correct. But if it is, that only means that, if possession conditions fully determine the intentional properties of a belief, there must be more to the identity of a belief than is determined by its intentional properties.[40] But then the Fregean has, as I said, the same problem the naïve theorist has.

The one theory that might avoid this conclusion is a radically holistic conceptual role semantics—radically holistic in the sense that the content of a belief is only individuated by *all* the inferential relations in which it stands. It is the radically holistic element of the view that gives this theory a chance: If absolutely all of the inferential connections matter, then, plausibly enough, there would be at least one such difference between any two of a thinker's beliefs.[41] For the same reason, however, content, so individuated, will almost never be shared. And the

[39] It's worth noting that this sort of response isn't at all available on Peacocke's new view (1998), according to which what individuates sense may involve principles that are not consciously but only tacitly known.

[40] Another option, in principle, would be to concede that the intentional properties of a belief are not determined by its content. But that is not in the spirit of Peacocke's position.

[41] Perhaps even this radically holistic view will not do: One can imagine certain sorts of symmetry, so that the beliefs would differ only in their inferential relations to one another. And then the contents differ only, so to speak, in so far as they are different.

problem with that result is not, as often seems to be supposed, that sense must be shared because sense is somehow essentially connected with language.[42] Nor is it simply that psychological laws will almost never have more than one instance, whence there will be almost no common explanations to be given of different people's behavior.[43] It is, rather, that intentional laws then will not have the sort of generallity we ordinarily suppose they do. One would have supposed, for example, that the explanation we give of why Fred became so upset when the grocer told him, "Sam Clemens has died", would still have applied even if, say, Fred had not believed that Clemens once lived in Missouri. Indeed, we find ourselves wanting to say: Fred would still have been upset even if he had not believed that Clemens once lived in Missouri *because* he still would have had the other mental states that caused him to become upset, such as the belief that Clemens was his neighbor. But on the radically holistic view, we can't say that. Fred would not then have had the same beliefs he now has; in particular, he would not have had the belief he now expresses as "Clemens is my eccentric neighbor" but, rather, a different belief he would then have expressed the same way. This is just a consequence of the fact that, on the radically holistic view, absolutely every inference in which a belief figures is partially determinative of its content. Psychological explanation would then not support the right sorts of counterfactuals. And that is why we do not want to radically holistic view.

3.2 Crypto-Fregean Views of Belief

I have argued that we cannot avoid the conclusion that there are two different mental states in which Fred might find himself, one that would constitute his believing (as we are putting it) that Mark Twain has died and one that would constitute his believing that Sam Clemens has died. The question is whether we can make this concession without also conceding that content must be individuated more finely than by Russellian propositions.

[42] Thus, the relation between the meanings of expressions and the contents of mental states is a dominant theme in John Perry's now classic paper "Frege on Demonstratives" (1993). More recently, William Taschek claims that an acceptable notion of sense must "support not only our usual assessment of the consistency, inconsistency, and the like of sentences as used by the same person, but also such assessments as they concern sentences used by different people" (Taschek, 1998, p. 330). My own view, for what it is worth, is that the notion of sense, for Frege, is primarily a *cognitive* notion and that it is a bold, and ultimately untenable, thesis that this same notion is fit to do serious work in the philosophy of language (Heck, 1995, 2002).

[43] Fodor, of course, has been making this point for decades (Fodor, 1987, p. 57). I remember recently seeing a paper specifically replying to it, but I can't now seem to find it. The point made in this reply was that, even if the law has only one instance, it could still be general in the sense that it *would* apply to any agent whose beliefs had the appropriate sorts of contents. But I think Fodor's point was always more the one that follows.

As I mentioned earlier (see page 8), few of those who have wanted to defend broadly Russellian accounts of belief-ascription have wanted to defend the naïve theory of belief itself. Rather, the common wisdom is that belief is a three-place relation between an agent, a proposition,[44] and something else. There is no agreement about what this third thing should be called let alone about what precisely it might be. Some writers regard it as something akin to a linguistic meaning—David Kaplan's (1978) character or John Perry's (1993) role, perhaps—and some regard it as a kind of representation, perhaps a sentence of natural language or a mental representation.[45] There does seem to be agreement, however, that adopting such a view is a way of avoiding a commitment to anything like Fregean senses. One thus regularly finds proponents of such views denying that the third term, whatever it is, is an aspect of content and, similarly, insisting that the content of a belief is simply a proposition. So we would do well to consider whether the naïve theorist might simply adopt some such view in response to Frege cases.

That depends. Let me first distinguish what I shall call 'crypto-Fregean' views from what I shall call 'representationalist' views. A ternary conception of belief is *representationalist* if the third term of the relation is some sort of representation that is individuated in wholly non-intensional terms. If, on the other hand, the third term is individuated at least partly in intensional terms, then the view is *crypto-Fregean*. What I want to do in this section is to justify the label 'crypto-Fregean': To adopt such a view is, contrary to what most of their defenders seem to intend, not to endorse an alternative to Frege's view that belief is a two-place relation between an agent and a 'thought' but rather to adopt a particular version of that view. Representationalist views will be discussed in the next section.

Let me speak, henceforth, of the third term of the relation as a 'way of thinking', just to settle terminology. So crypto-Fregeans hold that belief is a relation between a thinker, a proposition, and a way of thinking of that proposition.

The first point that needs to be made is that ways of thinking of *propositions* are not what are needed here (Heck, 1995, p. 80, fn. 5). Consider again the question why Fred comes to believe that his eccentric neighbor has died. Insisting that belief is not a binary but a ternary relation does put us in a position to resist the claim that, since Fred already at t_1 had beliefs with the contents that Sam Clemens had died and that Sam Clemens was his neighbor, he ought then to have been in a position

[44] I shall henceforth drop the modifier 'Russellian' and use the term 'proposition' just to mean: Russellian proposition.

[45] Many writers just aren't terribly clear about what the third term is supposed to be, which of course leaves them no worse off than Frege, whose remarks on modes of presentation are infamously cursory. And the distinction that I'm trying to make is rarely in the foreground, so people do not always locate themselves with respect to it. But Braun, Perry, and Salmon all seem to me to be crypto-Fregeans; Mark Richard (1990) is a clear representationalist.

to infer that his neighbor had died: We need only insist that the relevant inference could not be made from the latter belief and the belief that involved thinking of the proposition that Sam Clemens had died in way τ_1 but only from the latter belief and the belief that involved thinking of that proposition in way τ_2. But why *can* the inference be made in the second case? Why is τ_2 the right way and τ_1 the wrong way? The natural thing to say—indeed, the only thing to say, so far as I can see—is that the inference can be made when, and only when, Fred is thinking of *Sam Clemens* the same way both times. So it is not ways of thinking of propositions but ways of thinking of objects that we need.

The problem becomes more acute when we consider beliefs involving relations. Consider, for example, a belief with the content that Mark Twain was taller than George Orwell. How this belief interacts with other beliefs—say, one with the content that George Orwell was taller than Joyce Carol Oates—will depend upon how the agent thinks of both terms of these relations. No inference will be possible unless s'he thinks of Orwell the same way both times. And which belief results from the inference will depend upon how s'he thinks of Twain and Oates: S'he might arrive at the belief we can describe as the belief that Twain was taller than Oates; s'he might arrive at the belief that Clemens was taller than Rosamond Smith; and so forth.

To account for all the logical relations between beliefs involving singular representations, then, we shall need to regard the third term of the relation of belief not as a way of thinking of a proposition but as a collection of ways of thinking of the terms of the proposition. But then it is unclear why we should not simply regard belief as a *two*-term relation between agents and what we might call 'presented propositions', which can be taken to be complexes consisting of objects, properties, and the ways in which these entities are presented. But as Gareth Evans (1985, §VI) insisted, this view is very much in the spirit of Frege's. To be sure, no such view need incorporate all of Frege's views about sense. In particular, there is no need for a view of this sort to regard 'ways of thinking' as wholly independent of reference, as 'determining' reference in some strong sense.[46] But then, Evans himself was inclined to abandon these aspects of Frege's view, and he certainly thought himself a Fregean.

Just how close such a view is to Frege's will depend upon how one understands ways of thinking, in particular, upon how similar ways of thinking, so understood, are to Fregean senses. Fortunately, however, we do not need to resolve such questions for our purposes. The question here is whether the intentional properties

[46] Then again, Braun writes that "an agent stands in the believing or desiring relation to a proposition in virtue of standing in another psychological relation to an intermediary entity that determines the proposition that the agent believes or desires" (Braun, 2001, p. 256), so he actually seems to endorse this aspect of Frege's view.

of beliefs are individuated more finely than by Russellian propositions, where the intentional properties in question are all and only those that play a role in psychological explanation, and it is clear enough that crypto-Fregean views regard ways of thinking as playing just such a role. So crypto-Fregean views cannot help the naïve theorist.

3.3 Representationalist Theories of Belief

And so, again: We must allow that Fred's belief that Mark Twain has died and his belief that Sam Clemens has died are different beliefs. Can we nonetheless regard them as having the same content? The obstacle to our doing so, recall, is that we would then seem to be committed to the view that the psychological role played by a particular belief is not determined by its content. So the first question to ask is how our allowing that these two beliefs are different might help us explain why Fred becomes upset only when the grocer tells him "Sam Clemens has died". The answer must surely be that it is only the belief that Sam Clemens has died that is disposed to cause Fred to become upset, and it is is only when the grocer tells him "Sam Clemens has died" that he comes to have that belief. When Fred reads "Mark Twain has died", on the other hand, he comes to have the belief that Mark Twain has died, not the belief that Sam Clemens has died. Thus, these two states stand, as was suggested, in different relations to other of Fred's mental states.

Consider now the question why Fred did not already believe at t_1 that his neighbor had died. The answer will have to involve the fact that, just as the belief that Twain has died is different from the belief that Clemens has died, so the belief that Clemens is Fred's neighbor is different from the belief that Twain is Fred's neighbor. And though Fred did have both the belief that Twain had died and the belief that Clemens was his neighbor at t_1, he did not then have either the belief that Clemens had died or the belief that Twain was his neighbor. Why, though, should that be enough to explain why Fred was in no position at t_1 to infer that his neighbor had died? The language we are using to describe these beliefs may make such a conclusion seem natural, but it cannot license it.

How do things look from a Fregean point of view? Frege would have us say that Fred's Twain-beliefs involve one way of thinking of Twain and that his Clemens-beliefs involve a different way of thinking of Twain. But why does *that* allow us to explain why Fred is in no position at t_1 to infer that his eccentric neighbor had died? One might think the answer obvious. The inferences we are considering are, roughly, of the forms:

(1) $F(t)$; $c =$ the N; therefore, $F($the $N)$

(2) $F(c)$; $c =$ the N; therefore, $F($the $N)$

But the question, quite simply, is *why* (2) is the correct formalization only when the two beliefs involve the same way of thinking of Twain, (1) being the correct formalization if the two beliefs involve different ways of thinking of Twain. And the answer, I take it, is that, for Frege, this is simply one of the roles the notion of sense was supposed to play: For Frege, that is to say, sameness of sense is the standard by which we judge whether an argument equivocates. Or, to put the point in a way reminiscent of an earlier discussion (see section 2.3), sameness of sense is what is required if the premises and conclusion of an inference are to be related in the way they need to be for that inference to be permissible (Taschek, 1992; May, 2006; Heck and May, 2010).

It is important to see that only *identity and difference* of sense play a role in the Fregean story: The particular senses Fred associates with 'Twain' and 'Clemens' play no role at all. If this is not already obvious, consider this point: The senses Fred associates with 'Twain' and 'Clemens' could be swapped, and nothing in the foregoing would need to be changed; that is because the senses Fred associates with these expressions have not been identified except by using such descriptions as 'the sense Fred associates with "Twain"'. We don't know well enough how to talk about sense to do any better than that. So the main work the notion of sense is doing here is that it licenses us to treat Fred's beliefs that Twain has died and that Clemens has died as standing in different inferential relations with other of his beliefs.

Frege's own account of Frege cases thus has two parts. First, Frege claims that, because Fred thinks of Clemens under two different modes of presentation, he has two sets of beliefs about him that stand in different inferential relations with his other beliefs. For example, there are at least some inferential relations in which Fred's belief that Clemens is his neighbor stands to his belief that Clemens has died in which it does *not* stand to his belief that Twain has died. Second, Frege uses the fact that these beliefs stand in different inferential relations to explain the problematic aspects of Fred's psychology. The important point is that the notion of sense figures only in the *first* part of this explanation: Frege uses the fact that the beliefs have different contents to explain why Fred's belief that Clemens has died is related to his belief that Clemens is his neighbor in the way required for Fred to be able reasonably to infer that his neighbor has died. If we could offer a different explanation of why Fred's beliefs stand in the inferential relations they do, then, we would not need the notion of sense to solve Frege's puzzle. This fact, note, simply follows from the structure of Frege's own solution.

The thing to say here is *not* that Fred's belief that Twain has died is 'inferentially isolated' from his belief that Clemens is his neighbor. Fred is presumably quite capable of engaging in reasoning in which both beliefs play a role: For example, if Fred also believes, for whatever reason, that Clemens admired Twain, then

he might conclude that someone his neighbor admired has died. What we need to explain, rather, is why Fred is not prepared to infer from the premises that Twain has died and that Clemens is his neighbor to the conclusion that his neighbor has died, and why it would not be rational for him to do so, if he did. Similarly, we need to explain why Fred does not (and rationally need not) regard the belief that Twain has died as incompatible with the belief that Clemens has not died, though he does (and should) regard it as incompatible with the belief that Twain has not died, so that he can rationally believe both that Twain has died and that Clemens has not died but cannot rationally believe both that Twain has died and that Twain has not died.

As I have said, the notion of sense is designed by Frege precisely to play this role. It plays this role by both distinguishing and relating the contents of Fred's beliefs: It distinguishes the contents of his beliefs that Clemens has died and that Twain has died, and it relates the contents of his beliefs that Clemens has died and that Clemens is his neighbor. So what the naïve theorist needs is a way to distinguish and relate these beliefs without distinguishing their contents.

There is a familiar way to do this: Treat beliefs as having *logical forms* in something like the way sentences and formulae do; then explain how beliefs interact inferentially in terms of their having the logical forms they do. Indeed, one of the central motivations for the language of thought hypothesis is that inferential relations between mental states are most naturally explained on the assumption that mental states have syntactic properties: Logical relations among mental states become formal relations, and inference becomes computation (Fodor, 1975). The obvious way to implement the proposal we are discussing is thus to identify the logical form of a belief with the syntactic structure of the Mentalese sentence that expresses it. Belief could then be regarded as a relation between a subject, a proposition, and a sentence of Mentalese, and so we would have arrived at a representationalist form of the ternary view of belief.

In a sense, I am going to endorse this view: I am, that is to say, inclined to accept the language of thought hypothesis. But, on its own, I do not think that the view just outlined yields a solution to Frege's puzzle nor, as we shall see, that it can even play a significant role in the solution to Frege's puzzle. The problem is that explaining the cognitive changes that occur in Fred in terms of what sentences of Mentalese appear in his belief box seems to be incompatible with the thesis that psychological laws subsume psychological states in virtue of the intentional properties of those states. This is precisely the point, in fact, that drives Fodor to bite the bullet and deny that our behavior in Frege cases admits of any psychological explanation at all (Fodor, 1994, lecture 1).[47]

[47] My understanding of Frege cases is thus very close to Fodor's. But my characterization of

Suppose we were to explain Fred's coming to believe that his neighbor had died like this:

> Fred had a belief b_1 with the content that Clemens had died, that belief being expressed by the Mentalese sentence '$D(c)$'; Fred also had a belief b_2 with the content that Clemens was his neighbor, that belief being expressed by the Mentalese sentence '$c = \text{the } N$'; he then inferred the sentence '$D(\text{the } N)$', which in turn expressed his new belief that his neighbor had died.

If that is the right story, then, I submit, non-intentional properties of psychological states do indeed play an indispensable role in the explanation: The explanation makes explicit reference to particular Mentalese sentences, such as $D(c)$, and therefore lacks anything like the generality psychological laws formulated in terms of content were supposed to have. One can see this by considering what the corresponding psychological law would be: It too would make explicit reference to particular Mentalese sentences, and, as a matter of empirical fact, it would therefore probably have no other instances than Fred and so would lack the sort of generality we ordinarily suppose psychological laws to have. If everyone had the same language of thought—if the same sentence expressed the same content in my language of thought that it did in everyone else's—then that would be different, but we have no reason to suppose that is true.[48] Indeed, there is no obvious reason to suppose that the notion of 'same Mentalese sentence' so much as makes sense inter-personally. It is for largely this reason that, even if one does accept the language of thought hypothesis, psychological laws must still be stated at the level of content. The computational story in which Mentalese appears is supposed to be a story about how psychological states and processes are *implemented*. It follows that explicit reference to sentences of Mentalese should no more appear in psychological laws than does explicit reference to neurons.

We do not, however, have to make explicit reference to Mentalese sentences to explain why Fred came to believe that his neighbor had died. It is easy to see, first of all, that which Mentalese sentences express the beliefs in question is completely irrelevant to the explanation on offer. To make this explicit, we might quantify over such sentences, thus:

---

the problem does not depend, as Fodor's does, upon questions about how psychological states are implemented. That is, it does not depend upon the representational theory of the mind. The problem, as I see it, is more general.

[48] We have no reason to suppose, in fact, that there are *any* non-semantic properties shared by different people's representations of the same content. This is a persistent theme in Murat Aydede's writings on this topic (see Aydede, 2000a,b; Aydede and Robbins, 2001) and, in fact, is a point Fodor makes himself (Fodor, 1982, pp. 101–2).

> Fred had a belief b_1 with the content that Clemens had died, that belief being expressed by a Mentalese sentence *of the form* '$D(c)$'; Fred also had a belief b_2 with the content that Clemens was his neighbor, that belief being expressed by a Mentalese sentence *of the form* '$c =$ the N', where the use of the same term c here as previously signals that these beliefs share an element of their form; Fred then inferred the corresponding sentence *of the form* 'D(the N)', which in turn expresses his new belief that his neighbor has died.

This is a significant improvement: Even if everyone has his own language of thought, that will not prevent the corresponding psychological law from having plenty of instances. But we can go even further. How psychological states are implemented—for example, the fact, if it is one, that beliefs are computational relations to syntactically articulated mental representations—ought to be irrelevant to the explanation being given. What is relevant is the fact that Fred's beliefs stand in certain relations and not in others. How exactly a statement of these relations should enter psychological explanation is not a question I am now in any position to answer, but one option would be to say something like the following:

> Fred had a belief b_1 with the content <Clemens, having died>; Fred also had a belief b_2 with the content <Clemens, =, his neighbor>; these beliefs were 'formally related' *via* their respective first terms. Since these beliefs were 'formally related' in this way, Fred was then able to infer the belief with the content <his neighbor, having died>, where this belief is 'formally related' to b_1 *via* their second terms and to b_2 *via* their the first and last terms, respectively.

That Fred was in no position at t_1 to infer that his neighbor had died would then be explained as follows:

> At t_1, Fred had a belief b_0 with the content <Clemens, having died>; he also had a belief b_2 with the content <Clemens, =, his neighbor>. But these beliefs were not 'formally related' in any way and so Fred was in no position to infer any belief with the content <his neighbor, having died>.

The pattern is clear enough for present purposes.[49]

[49] A suggestion very close to this one has been developed by Taschek (1995; 1998) in connection with questions about belief-attribution. Our concern, of course, is with belief rather than belief-attribution, but reflection on Taschek's position was nonetheless important to the emergence of the view I am developing here. See, in particular, the discussion on p. 330 of "On Ascribing Beliefs" (Taschek, 1998), and see note 56 for some further remarks.

The term 'formally related' is a term of art. What it is supposed to mean is familiar from formal logic: Saying that Fred's belief that Clemens has died is 'formally related' to his belief that Clemens is his neighbor means that the beliefs have the feature we aim to capture in formal logic when we represent them this way: $D(c)$, $c = \text{the } N$; rather than this way: $D(t)$, $c = \text{the } N$. To put the point more intuitively, it means that an inference from these two beliefs that relied upon the identity of the subject of the first and the subject of the second would not depend for its correctness upon an additional premise asserting their identity: It would not equivocate, and it would not be enthymematic.

There can be no doubt that there are such inferences: inferences that *presume* rather than state the identity of objects mentioned at different points. If the inference

> Clemens has died
> Clemens is my neighbor
> So, my neighbor has died

were necessarily enthymematic, relying upon the unstated premise 'Clemens is Clemens', then the inference

> Clemens has died
> Clemens is my neighbor
> Clemens is Clemens
> So, my neighbor has died

would also be enthymematic, and the regress would have begun. (Perhaps the tortoise ought to have mentioned that to Achilles.) So there must be some such relation as the one I am calling 'formal'. What its nature is, is a question to which we shall return.

It is important to note that the presence or absence of these formal relations is no guarantee that a given argument will be valid: It is typically a *necessary* condition on the validity of an inference that its premises and conclusion should be formally related in a certain way; it is almost never a sufficient one. Indeed, the existence of such formal relations is every bit as important to our understanding of fallacious inferences as it is to our understanding of valid ones.[50] The following argument, for example, is fallacious:

(3) If Clemens has died, then Martha is upset.
Martha is upset.
So, Clemens has died.

[50] Thanks to Ernie Lepore for asking a question that prompted these thoughts.

But this argument is just incoherent:

(4) If Twain has died, then Martha is upset.
Martha is upset.
So, Clemens has died.

While it may not rational to believe the conclusion of (3) just because one believes its premises, one can nonetheless understand the mistake. But only seriously confused people would 'reason' as in (4). Yet the only difference between these arguments is that the first premise of (3) is formally related to its conclusion in a way that the first premise of (4) is *not* formally related to *its* conclusion.

We can now see what Frege's puzzle is actually about. It is widely appreciated nowadays that it doesn't really have anything to do with identity-statements. (That would be why identity-statements have hardly been mentioned to this point.) Frege himself does often use examples involving identity-statements when he introduces the puzzle, but he also uses other sorts of examples, such as the pair:[51]

Hesperus is a planet.
Phosphorous is a planet.

The usual understanding of the puzzle nowadays is thus that it somehow concerns substitution. But what our discussion reveals is that the puzzle really concerns something even more fundamental, namely: when thoughts (or sentences) are so related that a transition from some of them to another counts as *rational* even if *fallacious*; that is, the puzzle concerns what distinguishes a case like (3) from a case like (4). And if that is what the puzzle is about, then we can also understand why the puzzle was of such interest to Frege: It sits at the very foundation of logic.

As we saw, for Frege, the sorts of 'formal' relationships between mental states that we have been discussing are to be characterized in terms of sense. Suppose, for example, that Fred believes that Martha is upset and that, if Clemens has died, then Martha is upset, and that he concludes, as in (3), that Clemens has died. According to Frege, this inference is fallacious rather than incoherent because the thought that is the content of Fred's belief that Clemens has died is itself a constituent of—it is the antecedent of—the thought that is the content of his belief that, if Clemens has died, then Martha is upset. What I am calling the 'formal relationship' between the mental states therefore supervenes, for Frege, on an internal relation between the *contents* of those states. And that, indeed, is what is distinctive of Frege's view: Everything that is necessary for the evaluation of the correctness of an inference is present already in the *contents* of the mental states that are involved

[51] I have discussed Frege's own treatment of the puzzle elsewhere (Heck, 2003; Heck and May, 2006, 2010).

in that inference. For Frege, then, validity is in the first instance a relation between *contents*.[52]

The proposal I am making abandons this aspect of Frege's view. It implies that the correctness of an inference—a rational transition from one mental state to another—cannot be stated purely in terms of facts about the contents of the states involved. The formal relations that hold between mental states must also be specified, and the fact that certain beliefs do or do not stand in such formal relations does not supervene on those beliefs' contents. That, I am claiming, is what Frege cases show us. And for the very same reason, the proposal I am making implies that psychological explanations and laws cannot be stated purely in terms of facts about the contents of mental states, either. But, to emphasize, the alternative I am offering is *not* that we should also make reference to particular Mentalese sentences—that would be fatal—nor even that we should make some less direct use of the fact, if it is one, that beliefs are computational relations to Mentalese sentences. How psychological states and processes are implemented is, on my view, neither here nor there so far as psychological explanations and laws are concerned. What is important is whether particular beliefs stand in the sorts of relations I am calling 'formal' relations, and the foregoing is intended as a proof that there are such relations.

In the end, then, we do *not* need to be able to assign logical forms to beliefs to explain how people behave in Frege cases. We do not, that is to say, need to ascribe a *particular* logical form to Fred's belief that Clemens has died and a *particular* form to his belief that Clemens is his neighbor and then determine, by examining these forms, whether the two beliefs stand in an appropriate formal relation. It may well be that two beliefs' standing in such a relation is, as a matter of fact, ultimately to be explained in terms of facts about how cognitive states are implemented. But psychological explanation need advert only to the fact that the beliefs do or do not stand in certain formal relations.

So I am not denying the language of thought hypothesis. I am claiming, rather, that it is not needed for the solution of Frege's puzzle. As far as the resolution of Frege cases is concerned—and, perhaps, so far as intentional explanation and intentional laws, generally, are concerned—the sorts of formal relations among beliefs that must be mentioned may be treated as psychologically primitive: We can make reference *directly* to these relations in giving intentional explanations and in stating intentional laws.[53] Whether formal relations among beliefs are metaphysically primitive is a different question, and surely they are not. And of course there

[52] It may be worth emphasizing that one's views about the relation between inference and validity are irrelevant to the issues under discussion here. But it will become important later that we not run these together.

[53] Taschek (1998, p. 332) makes much the same point in connection with attribution.

are alternatives to the language of thought hypothesis. One might endorse a sort of inferentialism: The belief-states in which Fred might find himself constitute a kind of inferential network, and the place a particular state occupies in such a network is a purely extrinsic feature of that state, which might as well be regarded as having no internal structure whatsoever;[54] formal relatedness will then be a feature of the inferential network. But the metaphysics of formal relatedness is not at issue here. Indeed, the point is precisely that solving Frege's puzzle does not require us to resolve that issue. Since I myself lean strongly towards the language of thought hypothesis, however, I shall occasionally help myself to it, since doing so often simplifies the exposition.

3.4 Summary

Here, in short summary, is the dialectical progression to this point, as I understand it.

The Fregean: Frege cases show that Fred's believing that Clemens has died and his believing that Twain has died are distinct mental states. So they must have different contents.

The Naïve Theorist: I agree that they are different states. But why must they have different contents?

The Fregean: These states play different causal and explanatory roles, and psychological explanation is intentional explanation; psychological laws are intentional laws; and mental causation, if it is to be worth saving, must be causation in virtue of intentional features. That's why they must have different contents.

The Naïve Theorist: That isn't so much an argument as a challenge. And I think I can explain how the beliefs that Clemens has died and that Twain has died could play different roles in Fred's cognition without having different contents. They do so in virtue of their being differently connected, inferentially speaking, with other of Fred's mental states. Thus, the belief that Clemens has died—a certain token mental state with the content <Clemens, having died>—can get together with Fred's belief that Clemens is his neighbor to cause a new belief that Fred's neighbor has died. His belief that Twain has died—which is a certain other token mental state with the same content—can*not* get together with his belief that Clemens is his neighbor to cause a new belief that his neighbor has died. The difference is thus real enough. But it is due to the fact that different beliefs with the same content can stand in different 'formal relations' with other beliefs, not to the alleged fact that these beliefs have different contents.

[54] This issue has a long history in Frege scholarship where it appears in the form: Are thoughts complexes composed of senses (Heck and May, 2010)? As far as the substantive issue is concerned, this view is common among inferentialists, such as Robert Brandom (1994).

The Fregean: These 'formal relations', as you call them, cannot simply be brute: There must be some story to be told about when they hold and when they do not. I have such a story to tell: The inferential relations that hold between mental states are determined by the contents of those states, in particular, by the senses that appear in those contents. Quite obviously, you are barred from any such account: According to you, the inferential relations in which beliefs stand cannot supervene on content. But then it appears that something non-intentional must enter psychological explanation, for what grounds the formal relations are non-intentional facts.

The Naïve Theorist: I agree that it would be strange to regard the 'formal relations' as brute. So their obtaining or failing to obtain must supervene on something else. But it is no part of my view here to say what that is. It might be a lot of things. For what it's worth, I suspect that the language of thought hypothesis is true and that formal relations supervene on Mentalese syntax. But it just doesn't follow from the fact that formal relations are constituted by syntactic facts about Mentalese that there is no difference between appealing to formal relations between beliefs and appealing to syntactic facts about Mentalese. The most obvious difference is that explanations and laws formulated in the latter terms sacrifice generality in a way that explanations and laws formulated in the former terms do not. Ultimately, the point is really quite a familiar one: We're dealing with different levels of explanation. Psychological explanation appeals to formal relations between mental states, relations that are *implemented* by syntactic facts about Mentalese. And just as psychological explanation is distinct from neurological explanation, so it is distinct from any explanation that would appeal to syntactic facts about Mentalese.

The Fregean: Frankly, I don't find that very satisfying. It seems to me as if there is some sense in which syntax is what's really doing the work on your view—some sense in which Fred's behavior isn't really being explained in terms of what he believes, in terms of content.

The Naïve Theorist: Well, frankly, I don't always find the view satisfying myself. So why don't you see if you can't get at what's bothering you some other way?

4 Objections and Replies

4.1 Attribution and Communication

Objection There is a strong intuition that:

(5) If Fred believes that Clemens has died and also believes that Clemens is his neighbor, then, *ceteris paribus*, he will be in a position to infer that his

neighbor has died.

But according to you, there is no such law.

Reply That depends. I certainly do deny this:

(6) If Fred has a belief with the content <Clemens, having died> and another with the content <Clemens, =, his neighbor>, then, *ceteris paribus*, he will be in a position to infer a belief with the content <his neighbor, having died>.

In my view, and contrary to the view of Fodor's we discussed in section 1, Frege cases show that there is no such law. But there is no immediate conflict between the intuition expressed by (5) and my denial of (6), as even a momentary glance at the two statements shows. To get a conflict, you have to assume that I am committed to the view that (5) entails (6). Indeed, what underlies the objection is presumably the thought that, on my view, (5) and (6) must express the same thing, because, on my view, "*N* believes that *S*" is true if, and only if, *N* has a belief whose content is the Russellian proposition that is expressed by the sentence *S*. That is: The objector is supposing that I am committed to a naïve Russellian treatment of belief-attributions.[55] But I have repeatedly insisted that the views about belief that I am defending here are wholly independent of any view about the semantics of belief-attributions. And, as it happens, I reject the naïve Russellian account for the simple reason that I think, as most people do, that "Fred believes that Clemens had died" may be false even if "Fred believes that Twain has died" is true. I think, moreover, that some utterances of "Fred believes that Clemens has died and also that Clemens is his eccentric neighbor" will be true only if the beliefs mentioned are formally related in the way indicated by the language used in ascribing them. It will be no surprise that I have no semantics for belief-attribution—or, better, for complement clauses—that will deliver this conclusion.[56] My point is simply that, if belief-attribution does work this way, then we need not deny the intuition the

[55] Part of the point here is that, in so far as (5) expresses something about which one can have intuitions that have some claim to be respected, (5) itself must be a statement that is expressed in ordinary language.

[56] There are several accounts in the literature that are broadly consistent with the view towards which I am gesturing. Taschek's account, mentioned earlier (in note 49), is very much in this spirit, and interpreted logical form accounts share something of the underlying idea, too (Larson and Ludlow, 1993); so too do views that rely upon co-indexing (Fiengo and May, 1994).

As I have referred to Taschek's view a few times now, let me say a few words about it. The proposal is that we should "reject the idea that semantic content is subject to unrestricted compositionality and accept instead a principle that requires... the preservation of global logical structure" (Taschek, 1998, p. 331), where sentences like (i) "Clemens has died" and (i′) "Twain has died" have different 'global' logical structures though they have the same 'local' logical structure. So, if sameness of content is guaranteed only if substitution preserves global logical structure, then (ii) "Fred

objector is expressing. It will be perfectly consistent with the view being defended here.

Objection When the grocer tells Fred "Sam Clemens has died", Fred forms a new belief, the one you have been describing as the belief that Sam Clemens has died.[57] But, according to you, this is a belief that has the same content as the belief that Mark Twain has died, and that is a belief Fred already had. Why does Fred form a new belief with the same content rather than simply regarding himself as having acquired additional evidence for the belief he already held?

More simply, consider t_1, when Fred was reading the paper and saw the obituary for "Mark Twain". Why did he form the belief that Mark Twain had died rather than the belief that Sam Clemens had died, if these beliefs have the same content?

Reply Exactly how this objection should be answered depends upon how we should understand the use of language in communication. But there is at least one view that permits an answer.

Suppose that one's occurrent understanding of an uttered sentence consists in one's knowing a T-sentence for it, one delivered by the operation of the language faculty (Heck, 2005, 2007b). Thus, to understand the sentence "snow is white" as uttered by a normal speaker of English is to know that the utterance in question is true if, and only if, snow is white. Then, just as the belief that Twain has died is different from the belief that Clemens has died, so the belief that

> (SC) "Sam Clemens has died" is true if, and only if, Sam Clemens has died

is different from the belief that

> (MT) "Sam Clemens has died" is true if, and only if, Mark Twain has died.

---

believes that Clemens has died" and (ii′) "Fred believes that Twain has died" are not guaranteed to have the same content and so may have different truth-values. But the form of compositionality Taschek endorses (what he calls GLS-compositionality) implies that even the *simple* sentences (i) and (i′) need not have the same meaning even if 'Clemens' and 'Twain' do, and it implies this for the very same reason it implies that (ii) and (ii′) need not have the same meaning. I should emphasize that GLS-compositionality does not actually tells us that the sentences in these pairs *do not* have different contents, only that they *need not*, so it is consistent with Taschek's view that only the sentences in the second pair have different meanings. But the real problem is that we lack any account of how, consistently with GLS-compositionality, the meaning of a complex expression is determined by the meanings of its parts. To put it differently: It is one thing to state a restricted form of compositionality; it is another to produce a semantic theory consistent with it.

[57] Note that, if the remarks in the previous reply are correct, this language may be perfectly in order and so not be used simply 'for lack of better language', as I said above.

When the grocer utters the sentence "Sam Clemens has died", Fred's linguistic competence issues in the occurrent belief (SC). Fred takes himself to have reasons to regard the grocer as speaking the literal truth, so he discharges the left-hand side and forms the belief that Clemens has died. Had his linguistic competence instead issued in the belief (MT), the parallel inference would have yielded that Twain had died, and in that case Fred would merely have acquried additional evidence for a belief he already held.

It should be obvious that a similar story can be told about why Fred formed the belief that Twain had died when he read in the paper "Mark Twain has died" rather than forming the belief that Clemens had died.

It is worth noting that, while these remarks obviously do not commit me to the distinction between sense and reference, they do something remarkably similar. According to this view, Fred understands the expressions 'Mark Twain' and 'Samuel Clemens' differently, and this difference makes its presence felt even in the most literal communication. The difference lies in the fact that Fred knows that 'Mark Twain' refers to Mark Twain but not that it refers to Sam Clemens (compare Heck, 1995; McDowell, 1998). And if one thinks, as I do, that semantic theory should be in the business of expounding the knowledge of meaning that constitutes a speaker's semantic competence,[58] then we may also say that there is a *semantic* difference between these two names: What a correct semantic theory for Fred's idiolect would report about the one name is different from what it would report about the other. In that sense, they mean different things to Fred.[59]

4.2 Psychological Explanation

Objection The proposal you are defending is that psychological explanation should make reference both to the contents of psychological states and to the inferential relations in which such states stand to one another. But surely this is an old idea. According to so-called two-factor theories, such as that defended by Ned Block (1986), there are two aspects to the content of a mental state: Its wide content, which we may identify with the Russellian proposition you are calling its content *simpliciter*, and its narrow content, which is the state's conceptual role. So it looks very much as if Block has precisely the resources you are deploying. What, then, is different about your view?[60]

[58] The prime mover in this tradition is James Higginbotham (1985; 1989; 1992). Other defenders of the view include Richard Larson and Gabriel Segal (1995, ch. 1).

[59] But surely, one might object, there is no objective sense in which these words have different meanings! True, but why suppose that 'meaning', in this 'objective' sense, has any role to play in the semantics of natural language?

[60] A similar worry is voiced by Akeel Bilgrami (1998), and the reply is much the same. Bilgrami

Reply The similarities between my view and Block's are illusory.

First, and most importantly, the sorts of formal relations to which I am claiming appeal must be made in psychological explanation are very different from the sorts of inferential relations on which conceptual role theorists have focused. The latter would include, for example, the inference from "It's a cat" to "It's an animal", and, as Block makes clear, it must include a fair bit besides if we're not to get the same narrow content for 'cat' that we get for 'dog' (Block, 1986, pp. 628ff). The formal relations to which the naïve theorist must appeal to handle Frege cases, on the other hand, are, well, *formal* relations that serve merely to distinguish beliefs Fred regards as, of their very essence, concerning the same object from beliefs he either does not regard as being about the same object at all or regards as being about the same object only because of collateral information he possesses. For this reason, Fred's Twain-beliefs and his Orwell-beliefs will stand in precisely the same sorts of formal relations to other beliefs, even though they will stand in these relations to different beliefs—his Twain-beliefs to other Twain-beliefs, and his Orwell-beliefs to other Orwell-beliefs. Similarly, his cat-beliefs and his dog-beliefs will stand in the same formal relations to other beliefs, though they will stand in these relations to different beliefs: the cat-beliefs to other cat-beliefs; the dog-beliefs to other dog-beliefs.

Second, and relatedly, although one sometimes encounters language that suggests otherwise, two-factor theorists do not *identify* narrow content with conceptual role. The conceptual role of a state is supposed to *determine* its narrow content. This is not an optional feature of such views. On two-factor views, psychological laws are at least as involved, and are usually much more involved, with narrow content than they are with wide content: It is in virtue of the fact that two agents have beliefs with the same *narrow* contents that they will instantiate the same psychological laws. So, if psychological explanation is to be intentional explanation, sharing narrow content must be sharing intentional properties. But it simply isn't obvious why sharing conceptual roles is sharing any sort of intentional—that is, representational—property. One therefore has to earn the right to this claim by explaining how conceptual role determines some intentional property we can identify with narrow content. That is why Block is so concerned to argue—though the details of his particular view are surely optional—that narrow content is a function

argues that syntactic facts concerning Mentalese depend upon what kinds of inferences get made. Some of his arguments seem to have more to do with how *we* might decide such questions rather than with what constitutes such facts. But even if we waive this point and suppose that syntactic facts are constituted by facts about what inferences get made, the kinds of inference that individuate syntax will be the broadly 'formal' inferences on which I've been focused, rather than the kinds of inferences that distinguish dog-beliefs from cat-beliefs.

from context to wide contents (Block, 1986, pp. 643ff).[61] It is also why Block identifies it as "*the* crucial question" for two-factor views "what counts as identity and difference of conceptual role" (Block, 1986, pp. 629). I hope it is obvious that no such questions arise for the view I am defending.

Objection The examples on which you have been focused largely concern the explanation of changes of mind—cognitive or affective—rather than the explanation of behavior. That, perhaps, is why the examples have featured proper names rather than demonstratives or indexicals. So consider the following sort of example. Fred is absolutely terrifed of aircraft carriers but now has the bad luck to find himself in a Perry-inspired philosophy example. So there he is looking out one window when he sees the stern of a ship, one he does not identify as an aircraft carrier, and then he turns and looks out another window, sees the bow of a ship, immediately does identify it as an aircraft carrier, and proceeds to jump out the first window, thus running away from where he takes the aircraft carrier to be. Please explain Fred's behavior.

Reply What this sort of example shows is that we need to extend our conception of the formal relations in which a belief might stand to involve relations to perceptual states. Demonstrative thoughts, I take it, are thoughts that are connected in some special way to perceptual representations (Evans, 1982, ch. 5). If one thinks that perceptual representations are fully conceptualized and so that the very same concepts that occur in thought can also occur in perception, then there is no problem whatsoever: The perceptual representation of the ship as seen through the second window simply involves the very same 'demonstrative concept' that is deployed in Fred's belief that $\textit{that}_{\text{bow}}$ ship is an aircraft carrier, and the story told about Twain and Clemens applies almost without change. More importantly, one can tell much the same story even if one thinks, as I do (Heck, 2000, 2007a), that perceptual representations are not fully conceptualized, so long as they are conceptualized in the relevant respect, that is, so long as there are object-representations of some sort in perception (Siegel, 2006). The perceptual representation and the demonstrative thought could then be formally connected *via* the object-representation contained in the former and the demonstrative concept contained in the latter.

Even if perceptual representation is wholly unconceptualized, however, it is surely a constraint on any decent view that some sense be made of the idea that there are certain sorts of thoughts—demonstrative thoughts—that one can have only because, and only in so far as, one is (or at least appears to be) in perceptual contact with an object.[62] The fact that Fred forms the belief he does—that

[61] A similar point could be made, obviously, about any form of two-dimensionalism.

[62] Note that this is a claim about certain *thoughts*, that is, about certain token mental states, *not* a

that$_{\text{bow}}$ ship is an aircraft carrier rather than the belief that *that*$_{\text{stern}}$ ship is an aircraft carrier—is therefore to be explained in terms of the fact that the perceptual representation he has of the ship as seen through the second window is 'formally' connected—that's probably not the right notion, but it will have to do—with *that*$_{\text{bow}}$-beliefs rather than with *that*$_{\text{stern}}$-beliefs, and so on and so forth. The story one would need to tell would then be far more complex than the one about Twain and Clemens, but it would follow broadly similar lines. The fact that Fred, having come to believe that *that*$_{\text{bow}}$ ship is an aircraft carrier, jumps out the first window, whereas he would have jumped out the second had he come to believe that *that*$_{\text{stern}}$ ship was an aircraft carrier, is to be explained in much the same way.

One might well say, then, that demonstrative thoughts, as a class, have a distinguishing feature, one thoughts in general do not have: Demonstrative thoughts, as I have been saying, are connected to perception (and maybe also to action) in a special way. And perhaps there are useful generalizations to be stated about the members of this class. If so, then, demonstrative thoughts, as a class, form a psychological kind. But it is no part of my view, and I see no reason why it should be part of anyone's view, that all psychological kinds must have their boundaries fixed by content.

Objection You didn't say anything about indexicals. What, then, about self-conscious, first-person beliefs? More generally, what about so-called self-locating beliefs?

Reply Frankly, I'd like to declare myself out of space, because I don't really know what to say about indexicals. It seems clear that thoughts about myself, about here, and about now play a fundamental role in human cognition, and some account needs to be given of that fact. Moreover, it seems obvious that there are psychological laws about such thoughts: The effects my self-conscious thoughts have on my behavior are similar to the effects your self-conscious thoughts have on your behavior.

Now, as I just said, I see no reason that there should not be psychological kinds whose instances are unified by something other than a shared aspect of their content. So perhaps the thing to say is that self-conscious thoughts, and indexical thoughts more generally, are another example of such a kind. Even if that is so, however, something still needs to be said about what unifies this kind if it is not an aspect of content. It'd be nice if indexical thoughts could be fit into the mold of demonstrative thoughts: Perhaps the similarities between your self-conscious

---

claim about the contents of those thoughts. The claim is not that there are contents one can entertain only if one is in perceptual content with a particular object, but rather that there are certain cognitive states one can only then be in. There is therefore no commitment here to any sort of conceptual role semantics or, as Fodor has been calling it, 'conceptual pragmatism'.

thoughts and mine could be explained in terms of similar connections between those thoughts and certain sorts of perceptual, kinaesthetic, etc., states in which we might find ourselves. I'm in no position to say. I'm also in no position to say whether, if we had such an account, we would find ourselves tempted by the claim that the content of Twain's self-conscious thought that he is an author differs from that of his thought that Twain is an author. But even if the question were decided Frege's way, that would be but a small victory for him, since there is no prospect, so far as I can see, of parlaying that victory into a larger one.

Note, however, that this issue does not really have anything to do with Frege cases, and my claim here has been simply that Frege cases, by themselves, do not motivate a distinction between sense and reference. Indexicality has much more to do with Twin cases, about which a little more shortly.

Objection You argued earlier that your view is consistent with the thesis that psychological explanation is intentional explanation. The reason was supposed to be that the only reference to non-intentional features of mental states that is required in psychological explanation is reference to formal relations between states. But there is another worry, namely, that intentional content, on your view, plays no significant role at all. Consider Fred's buddy Barney. At t_0, Barney has heard nothing about anyone dying. Now suppose that at t_1 Barney comes to have a belief with the content: <Twain, having died>. The belief in question might be the one you have been describing as the belief that Twain has died; or it might be the belief that Clemens has died; or it might be based upon a demonstrative presentation of Twain. The supposition that Barney has a belief with the content <Twain, having died> does not decide among the many different beliefs with that content he might have. But then it seems as if the hypothesis that Barney has acquired a belief with that content has no determinate consequences whatsoever, and that strongly suggests that the contents of mental states are not, on your view, doing any real work.

Reply I fully appreciate this kind of worry. Indeed, when I first started working on this paper, I believed that considerations of this kind would ultimately serve to reveal why we *do* need some notion of sense. And, to be honest, I would not be shocked if, in the end, Frege won this battle on precisely this ground. But it now seems to me that this objection can be answered.

We need to distinguish two questions. One is the very general question what role, if any, content plays in psychological explanation or, relatedly, to what extent the semantic properties of mental states are causally efficacious. Some people have of course held that content is causually epiphenomenal (e.g., Segal and Sober, 1991). Some have even held that the representational theory of the mind has no need for any notion of content, since all psychological explanation can proceed

directly in terms of syntax (e.g., Stich, 1983). I doubt it, but I do not claim to know precisely what to say here. These are very hard problems.

Fortunately, however, they are not our problems. The relevant question here is whether giving the sort of explanation I have proposed we should give in Frege cases somehow *undermines* the view that content has some substantial role to play in explanation, mental causation, or what have you. And to that question, it seems to me, the answer is clearly "No". I am most certainly not proposing that Fred's changes of mind can be explained wholly in terms of the syntax of his mental states. On the contrary, although I am proposing that Fred's changes of mind should be explained, in part, in terms of formal relations between his mental states, the explanations as I stated them also make explicit reference to the contents of those states. Perhaps the reference to content is merely apparent; perhaps it can be explained away. But that is a different question, and nothing I have said here makes it any more or less likely that it can be.

Objection That may be, but it doesn't address the underlying worry. According to you, Barney's beliefs that Clemens had died, that Twain has died, and that *that* guy has died all have the same content. But then any psychological law that applied to one of them would also have to apply to the others, and that is absurd.[63]

Reply If I may first stop to pick a nit: It is not true that my view implies that any psychological law that subsumes one state with the content <Twain, having died> will also subsume all other such states. I suggested, in response to a previous objection, that demonstrative thoughts might form a psychological kind. If so, there might be laws that subsumed demonstrative thoughts with this content that did not subsume all such thoughts.

But, of course, the worry remains: Any law that subsumes Barney's belief that Clemens has died in virtue of its content also subsumes his belief that Twain has died. This is correct, but, so far as I can see, it is not a bug but a feature. One such law might be this one:

> If one has a belief with the content <Twain, having died> and one also has a belief with the content <Twain, =, Fred's neighbor>, where these two beliefs are formally related *via* their first components, then one will, *ceteris paribus*, be in a position to acquire a new belief with the content <Fred's neighbor, having died>, where this belief is formally related to the first *via* their first components and to the second *via* their first and third components, respectively.

This is the kind of law that I am proposing is at work in Frege cases, and it both

[63] Jacob Beck (2008) expresses a version of this worry.

does *and should* subsume both Barney's Twain-beliefs and his Clemens-beliefs. Nothing untoward follows: Suppose Barney has the belief that Clemens is Fred's eccentric neighbor but not the belief that Twain is. Then he will not instantiate the law if he acquires the belief that Twain has died but will if he acquires the belief that Clemens has died. If he did have the belief that Twain was Fred's eccentric neighbor, the law mentioned would predict that, *ceteris paribus*, Barney would be in a position to acquire the belief that Fred's neighbor has died if he acquired the belief that Twain had died. And all of that is correct.

So yes, the mere fact that Barney has a belief with the content <Twain, having died> tells us very little about how he is likely to behave, think, or feel. But that is no surprise. No single belief in isolation has any particular connection to behavior. How Barney is likely to behave given that he has a certain belief depends upon what else he believes, what he wants, and so forth.[64] That is old news, and it has been said in this connection before (Fodor, 2003, pp. 105–6).[65] My point is just that how Barney behaves also depends upon how his mental states are formally related to one another—and that invoking such formal relations does not undermine the explanatory ambitions of intentional psychology.

Objection You earlier quoted a remark from Block to the effect that it's essential that we figure out how to type conceptual roles. Part of Block's point is that we need to know how to type beliefs inter-personally if there are to be any psychological laws worth stating. But your view, I take it, is that there is no way to type beliefs inter-personally except in terms of their content. Fred's belief that Phosphorous is a planet is of the same psychological type as Barney's belief that Phosphorous is a planet and is also of the same type as Barney's belief that Hesperus is a planet. Surely there is something odd about that.[66]

Reply It is true that, as far as their contents are concerned, the beliefs mentioned are of the same type. But if the worry is that there will be no psychological laws that apply to the beliefs both Fred and Barney would express as "Hesperus is a planet" that do not also apply to the beliefs they would express as "Phosphorous is a planet", it can be answered in the same way previous objections of this kind have already been answered. Fred and Barney are in other respects cognitively similar. Both of them, for example, might be inclined, when they acquire the belief they would express as "Hesperus is a planet", to make a simple inference and thus to acquire the belief they would express as "The brightest object in the evening sky is a planet". That is because they both have the belief they would express as

[64] It is this insight that is behind the proposals we discussed in section 2.1. But it was not properly implemented there.

[65] Indeed, the point really goes back to Putnam's criticisms of behaviorism (Putnam, 1975).

[66] Something like this objection is pushed by Bradley Rives (2009, §3).

"Hesperus is the brightest object in the evening sky", and because these beliefs are formally related to one another in the obvious ways.[67]

The mere fact that Fred and Barney both have a belief they could express that way does not show that those beliefs have a content more fine-grained than < Venus, being a planet >. Frege, of course, wanted to regard the shared belief that Hesperus is the brightest object in the evening sky as partly determinative of the content of the singular concept *Hesperus*. Now it would be wrong to saddle Fregeans—indeed, wrong even to saddle Frege (Dummett, 1981, ch. 5)—with the description theory of names. The general idea is simply that the content of a singular concept is determined by certain of the beliefs in which it is deployed or by certain of the inferences in which it is involved. But, as Fodor (1998) has emphasized and as Quine (1953) was the first to argue, the problem is to say *which* such beliefs serve to determine the finer-grained content and which do not, and no-one, so far as I know, has an adequate answer to that question. Or rather: The only acceptable answers seem to be "All" and "None", and the former leads to a form of semantic holism that has no hope of answering the sort of objection we are considering.[68]

Objection But if beliefs are typed inter-personally only by content, then you are committed to regarding Fred and twin-Fred as sharing no singular beliefs. It will then be a challenge to explain the many commonalities in their behavior. That is: You have a problem with Twin Earth.

Reply My response to this objection is similar to my response to the last one. I would endorse what I take to be a fairly common idea nowadays,[69] namely, that, although Fred and twin-Fred do not share their singular beliefs, there are nonetheless many *non*-singular beliefs that they do share, and their common behavior can be explained in terms of their sharing those non-singular beliefs. The crucial point is that this 'common behavior' has to be described in *general* rather than singular terms: When the behavior is described in singular terms, there is no commonality to be explained. And if the behavior is described in general terms, then it is plausible enough that it can also be explained in terms of general beliefs, and Fred and twin-Fred share lots of general beliefs.

There are all kinds of problems with this view (Segal, 2000b), but my focus

[67] Admittedly, there is a kind of awkwardness here. I can't say that Fred and Barney both believe that Hesperus is the brightest object in the evening sky, at least, not without relying more than I should upon the language we use to describe those beliefs. Note, however, that it does not follow that "Fred and Barney both believe that Twain has died" will be true so long as both Fred and Barney have a belief with the content < Twain, having died >. To think it did would again be to confuse belief with belief-attribution.

[68] Since no two people share all their beliefs, no two people will share any of their beliefs.

[69] The idea has its origin, for me, in Evans (1985). Peacocke (1993) developed it in ways that made it compelling.

here hasn't been on Twin Earth. I've just been trying to argue that Frege cases don't refute the view that psychological content is Russellian. So, yes, it may be that the need to explain the sorts of commonalities mentioned will require some notion of narrow content. I doubt it, but it's a different issue.

4.3 Semantics and Validity

Objection You have been speaking throughout of the contents of beliefs as Russellian propositions. But why not simply regard them as sets of possible worlds? Could one not defend that view in essentially the same way you have defended your favored view, by invoking formal relations between mental states? Worse: Couldn't the psycho-Fregean view be defended in much the same way you have defended the naïve theory? We could regard the content of a belief as being its truth-value and deal with the objections to that view the same way you deal with Frege cases: By appealing to formal relations between belief-states.[70]

Reply It would be no comfort to the Fregean if I were wrong and the friends of possible worlds were right, so, from our present point of view, that is a sectarian dispute. It *would* be a problem if the psycho-Fregean view could be defended along similar lines: That would seriously call into doubt whether content plays any role in psychological explanation.

But the psycho-Fregean cannot mimic the account I've offered. Consider, for example, this explanation:

> Fred had a belief b_1 with the content <Clemens, having died>; Fred also had a belief b_2 with the content <Clemens, =, his neighbor>; these beliefs were formally related *via* their respective first terms. He was therefore able to infer the belief with the content <his neighbor, having died>, where this belief is formally related to b_1 *via* their second terms and to b_2 *via* their the first and last terms, respectively.

If we try to adapt this pattern on behalf of the psycho-Fregean, we get something like the following:

> Fred had a belief b_1 with the content Falsity;[71] Fred also had a belief b_2 with the content Truth; these beliefs were formally related *via*... what? He was therefore able to infer a belief with the content Falsity, where this belief is formally related to b_1 *via*... what? and to b_2 *via*... what?

[70] Fine (2010, p. 482) presses a similar objection against an appeal Soames (2010, pp. 472–73) makes to logical form.

[71] For present purposes, we assume that the rumors of Clemens's demise were greatly exaggerated.

The problem is that there is not enough structure in the 'contents' of these beliefs for us to state how they are formally related. The really crucial point here is one made earlier. We could, of course, use the sentences "Clemens has died", "Clemens is Fred's neighbor", and "Fred's neighbor has died" to identify the contents of Fred's beliefs. For example, if we were to suppose with Frege that sentences name their truth-values, then we could state the explanation as follows:

> Fred had a belief b_1 with the content Clemens has died (i.e., the content snow is green); Fred also had a belief b_2 with the content Clemens is Fred's neighbor (i.e., the content pigs are mammals). These beliefs were formally related in ways suggested by the language used in stating their contents, and so Fred was able to infer a belief with the content Fred's neighbor has died (i.e., grass is pink), which was related to b_1 and b_2 in ways again suggested by the language used in stating their contents.

The difficulty is that the reference to 'the language used in stating their contents' makes the theory overly sensitive to the language used in stating the law, as the parenthetical re-statements make clear. If the content Clemens has died *is* the content snow is green, then substitution of the one of these phrases for the other cannot make a difference to the truth of the law being stated,[72] and yet it clearly would, since it actually makes it unintelligible. The contrast with my view should be clear: On my view, beliefs are inferentially related *with respect to aspects of their content*, and what relations do or do not obtain can be, and has been, stated in ways that are independent of the language used in saying what those contents are.

It should be clear that a similar objection applies to the possible worlds approach. On that view, the contents of beliefs are certain sets, and the formal relations that obtain between different beliefs cannot be stated in terms of relations between those sets, since the sets too lack the requisite structure. I'll leave development of the point as an exercise.

Objection The formal relations between beliefs can be stated in other terms. Indeed, these other terms are obvious. Beliefs need only be regarded as having 'global logical forms', in Taschek's (1998) sense, and then the formal relations between beliefs can be stated in terms of them. Fred's belief b_1, for example, can be regarded as having the logical form $D(c)$; his belief b_2, the logical form $c = \text{the } N$; and then the beliefs are related in virtue of the fact that they share an element of their logical forms. So the psycho-Fregean, for example, could say:

[72] Unless, of course, these phrases occur in intensional contexts. But these phrases are supposed to be *referring* to the contents of Fred's beliefs, not expressing them, whence the context is not intensional but extensional.

> Fred had a belief b_1 with content Falsity and the logical form $D(c)$; a belief b_2 with content Truth and logical form $c =$ the N; these beliefs are formally related in virtue of the shared element of their logical form; so Fred was able to infer a belief with content Falsity and the logical form $D(\text{the } N)$.

Why doesn't that work?

Reply Because the reference to the *content* of Fred's belief, in the last clause of the explanation, is completely gratuituous. In this particular case, the fact that Fred's new belief is false can indeed be inferred from what we know about the truth-values of his prior beliefs: If $D(c)$ is false and $c =$ the N is true, then $D(\text{the } N)$ will be false. But that is a special case. If $c =$ the N is false, then the truth-value of $D(\text{the } N)$ is independent of that of $D(c)$.

So change the example slightly. Suppose that Barney's belief b_2 was not true but false. Then, while we can still explain why Barney was in a position to infer a belief with the logical form $D(\text{the } N)$, we are in no position to decide what the truth-value of this belief might be. The psycho-Fregean therefore has no way to answer the question why Barney was in a position to form the belief he did: one with content Truth and the logical form $D(\text{the } N)$, and this is so even if content is understood as the psycho-Fregean would have us understand it.

Similar problems arise for the possible worlds theorist. The problem in this case is that the set of worlds in which $D(\text{the } N)$ is true is not determined by the set of worlds in which $D(c)$ is true and the set of worlds in which $c =$ the N is true. So the possible worlds theorist cannot answer the question why Barney formed the belief he did: one whose content was a certain set of worlds and whose logical form was $D(\text{the } N)$, and this is so even if content is understood as the possible worlds theorist would have us understand it.[73] So the possible worlds theorist cannot simply take over the solution I've given on behalf of the naïve theorist.[74]

[73] It's crucial that the inference involves something besides propositional inference. Those sorts of inferences can be explained on the possible worlds view, because sets of possible worlds form a Boolean algebra. The technical point amounts simply to the observation that sets of worlds to do not form a so-called cylindrical algebra—these are usually credited to Tarski (Monk, 1986, pp. 902ff)—that being what one would need to explain quantificational inferences.

[74] There is an objection that can be (and has been) made to my claim that different sorts of mental states may have different sorts of content (Heck, 2007a) that is entirely parallel to the objection we have been discussing. As I am about to explain, my view is that the sort of content a state has reflects the sorts of inferential (or, more generally, cognitive) processes in which it does and does not participate. The advantage to this view is that, if content is individuated in this way, then psychological explanations can be framed entirely in terms of the contents of psychological states—except, of course, for appeal to analogues of the present notion of 'formal relations'. The alternative is to say that cognitive maps, for example, do not have a special 'topographic' sort of content but simply that they have sets of worlds as their contents. Psychological explanations in which such

Still, that is merely negative, and one might yet want to know how we should decide what kind of content beliefs have. In fact, there is a more general issue here: how we should decide what kind of content to ascribe to states of any given kind. I've discussed this question in some detail elsewhere (Heck, 2007a), however, and do not propose to discuss it again here. But let me say quickly what I take to be the reasons that the contents of beliefs should be individuated at least as finely as Russellian propositions.

First, the fact that belief-states stand in certain formal relations to one another (and not in others) is essential to our ability to explain what needs explaining in Frege cases. I therefore take us to be licensed to regard these states as being logically articulated in a sense that should be uncontroversial. I am not saying that we must regard beliefs as having 'particular' logical forms, in the sense discussed earlier, let alone that we must endorse the language of thought hypothesis.[75] Ascribing logical form in this sense is simply a way of systematizing the formal relations in which belief-states stand to one another. For example, saying that Fred's belief that Twain has died has the logical form $\phi(\alpha)$[76] is a way of characterizing the formal relations in which it stands to other beliefs. These are the same sorts of formal relations in which his belief that Clemens has danced stands, and that is why it too is of the form $\phi(\alpha)$.

Second, this logical articulation is not merely syntactic but also has a semantic aspect: Other beliefs that are formally related to Fred's belief that Twain has died *via* its α-component (if I may put it that way) share an intentional feature with it, namely, that they too are about Twain. And this shared intentional feature is implicated in at least some of the explanations in which these beliefs are implicated. For example, the fact that Fred's belief that Twain has died is about Twain might be involved in explanations of why Fred acts toward Twain (or Twain's body) in certain ways.[77]

If so, then psychological explanations that mention Fred's belief that Twain has died appeal to intentional features of this belief that are determined neither by its

states figure will then appeal not just to content but also to features of the underlying representations. But this objection can be answered in much the same way as in the text. Suppose given a map-like representation with a certain set of worlds as its content. Now suppose a particular marker is moved from one place on the map to another. There is simply no way to predict, given just these meager resources, what the content of the altered map will be, and so my opponent will be unable to explain why the rat now thinks the food is behind it.

[75] Martin Davies (1992; 1998) has argued, however, that a weak form of the language of thought hypothesis may well follow.

[76] Of course, the actual form will be more complicated, in virtue of their being many more formal relations than this simple form indicates, but the complications do not matter here.

[77] The same kind of thing can be said about the ϕ-component: That Fred's belief that Twain has died is about dying is implicated in at least some explanations involving it.

truth-value nor by the set of worlds in which it is true. That, it seems to me, is enough to show that its content is not just a truth-value or a set of worlds: The intentional features of the belief simply outstrip both its truth-value and the set of worlds in which it is true. Russellian propositions are designed precisely to remedy this problem: A Russellian proposition is just an encoding of the intentional features we have been discussing, since it incorporates both the logical structure of the belief and the intentional features corresponding to the elements in that structure; that is, it encodes both the syntax of the state and the semantics of the constituents. I suppose we could say, if someone absolutely insisted we do so, that the 'content' of the belief was just a set of worlds while also saying that the belief had intentional features not determined by that set of worlds. But that would just be a crypto-Russellian view, one that used the word 'content' in a way every bit as idiosyncratic as the way crypto-Fregeans use it.

Now, to be sure, I have said nothing here to defend either of the two claims on which this defense of the Russellian view depends. These claims, again, are: (i) that the formal relations among beliefs are sufficiently robust to entitle us to regard beliefs as having logical structure; and (ii) that this articulation is not merely syntactic but has an explanatorily relevant semantic aspect. I take these claims to be pretty plausible. But what matters in the present context is the fact that the *Fregean* is in no position to object to either of them. In particular, if there are no explanatory purposes for which it is essential to invoke intentional features of beliefs not determined by the set of worlds in which they are true—in particular, to invoke semantic properties of constituents of its logical form—then the Fregean is in even worse trouble than the Russellian.

Objection Validity is fundamentally a semantic notion: Whatever *A* and *B* may be—contents, sentences, mental representations, or what have you—whether *A* entails *B* should not depend upon anything but their semantic properties. And on Frege's view, this is so: The thought that Clemens has died does not entail the thought that Twain has died, and so it is clear enough why Fred ought not infer the one from the other. On your view, though, the proposition that Twain has died *does* entail the proposition that Clemens has died, trivially, and so it is entirely unclear why Fred ought not infer the one from the other. You can say, if you like, that Fred's (token) belief that Twain has died does not entail his (token) belief that Clemens has died, but then entailment is not a purely semantic notion, and that is unacceptable.

Reply This objection confuses validity with inference.[78] It is *not* in general

[78] Precisely how we should understand the relation between validity and inference is a difficult question. Gilbert Harman (1988) has claimed that the one has essentially nothing to do with the other. I think Harman's view more extreme than necessary, but the distinction needs respecting.

true, on anyone's view, that, if *A* entails *B*, then it is rational to infer *B* from *A*. For example, Fermat's Last Theorem is provable in so-called von Neumann–Bernays–Gödel set theory,[79] which is axiomatizable by a single sentence. Let this sentence be *NBG*, and let *FLT* be Fermat's Last Theorem. Then *NBG* entails *FLT*, but prior to Wiles's proof it would have been insane for someone who believed *NBG* simply to have inferred *FLT*. It is a valid inference, but that does not make it a *rational* one.

Given Wiles's proof, of course, one can make such an inference, but then one is not just inferring *FLT* from *NBG* but from *FLT* and the claim that *FLT* if *NBG*, which is what Wiles's proof establishes. We can all agree that this inference is rational, but it cannot be the general case: That suggestion leads directly to the regress the Tortoise uses to slow down Achilles (Carroll, 1895). So it is a nice question what else is required of an inference, besides its validity, if it is to be rational. I obviously cannot answer this much discussed question here (or elsewhere, for that matter).[80] But it is a natural idea that the reasonableness of 'primitive' inferences requires not just that they should be valid but also that they should satisfy certain formal conditions. For example, it is widely agreed that inference by *modus ponens* counts as reasonable, and whether an inference is an instance of *modus ponens* is determined by formal (that is, syntactic) properties of its premises and conclusion. Why satisfaction of such formal conditions is necessary, and whether it is sufficient—those are the difficult questions. But they are not our questions here. What matters for present purposes is simply that there is nothing at all novel about the suggestion that the notion of *rational inference* has a syntactic component, and I need make no stronger claim than that.

This view does differ from Frege's. Indeed, I am tempted to suggest that it is here that we find bedrock: What most fundamentally distinguishes Frege's view from mine is that he regards the validity of an inference as completely determined by the contents of the mental states involved in the inference, whereas I do not. I am further tempted to suggest that Frege would have regarded my view as unacceptably psychologistic: Logic, he always insisted, must not concern itself in any way with mental states. But my view is not psychologistic. Logic concerns itself with whether an inference of a specified form is or is not valid, and that question has nothing to do with anyone's mental states. But it simply does not follow that the validity of a particular inference—a particular transition between mental states—is

[79] Nothing nearly so strong as NBG is needed here, but it is has the advantage that it is relatively familiar. Colin McLarty has recently shown that the proof of FLT can be formalized in a theory that has the same strength as simple type theory, and it is widely believed that it is in fact provable in Peano arithmetic.

[80] The exchange between Paul Boghossian (2003)and Timothy Williamson (2003) is a good place to start.

determined simply by the contents of the states involved. It would follow if the contents of the premises and conclusion determined the *form* of the inference, but that is precisely what I have been at pains to deny.

In this respect, the 'semantic relationism' of Kit Fine (2007) seems closer to Frege's view. Fine seems to agree with Frege that the correctness of an inference ought to be determined by a relation between the contents involved. If contents are Russellian, however, then the relation in question cannot supervene on intrinsic properties of the contents, so there must be some extrinsic relation involved, as well. On my view, by contrast, the correctness of an inference is *not* determined by any relation just between contents but by a relation between *representations* of those contents. But, obviously, Fine and I are in many ways thinking along the same lines, and, technically speaking, the two frameworks are likely to be inter-translatable. Nonetheless, the significance of the disagreement between us should not be underestimated. It goes very deep indeed. My view is as it is because I do not believe that the notion of content is intelligible absent some notion of a representation that has that content: something that also has *non*-semantic properties, such as a sentence or a mental state. As a slogan: No representation without representations. Someone who disagrees will no doubt want to explain inference directly in terms of propositions, and, to them, it will therefore seem as if the approach taken here appeals to something inessential.

5 Closing

Frege's notion of sense is multi-faceted: Sense is the content of propositional attitudes; it is what determines reference; it is indirect reference; it is what is grasped when we understand an expression; and it is much more. It is a common observation nowadays that these various roles are, at least *prima facie*, in tension with one another, and anyone who wants to develop a broadly Fregean notion of sense must decide which of these roles to preserve and which to discard.

I have argued here that the notion of sense is not needed for the solution of Frege's puzzle. It does not follow that the notion of sense is not needed. What does follow is that the notion of sense cannot be motivated entirely by Frege's puzzle, that is, that one cannot establish a need for the notion of sense while focusing exclusively upon its role as the content of attitudes. That this role is, at the very least, primary has been a persistent assumption in the work of some self-described Fregeans. It is, I can say with some authority, an assumption that is very much in the foreground of my own discussions of sense (Heck, 1995, 2002), and this orientation is one I take myself to have inherited from Evans. Evans does not regard sense as determining reference in any but a trivial sense (see, for example, Evans,

1982, §4.2); he hardly mentions the question of indirect reference; and questions about linguistic meaning play but an insignificant role in *Varieties of Reference*.[81] Rather, for Evans, sense is first and foremost the content of attitudes, and Frege's puzzle is what establishes the need for it. What other roles it might play is an open question.

There is another tradition, however. As mentioned earlier, Peacocke characterizes sense in terms of what he calls 'possession conditions': The content of a concept, on his view, is determined by what is required of a thinker if s'he is to grasp that concept. Peacocke argues that sense, so characterized, can serve as psychological content, but he insists equally strongly on the thesis that sense determines reference: Each possession condition must be paired with a 'determination theory' that explains how the semantic value—the reference, in a generalized sense—of the concept is determined by its possession condition (Peacocke, 1992, §1.3). For example, the semantic value of the concept of disjunction is supposed to be fixed by its possession condition in virtue of the fact that there is only one (classical) truth-function that validates the introduction- and elimination-rules for disjunction.[82] More recently, Peacocke has emphsized that, on his view, the rationality of an inferential transition "is to be philosophically explained in terms of the nature of the intentional contents and states involved in the transition" (Peacocke, 2004, p. 52). So the sense associated with a given concept serves also to explain the rationality of the inferences involving it, and, Peacocke argues, it does so in a way that reveals that these inferences are fundamentally *a priori*.

I noted earlier that, so far as Frege's solution to his puzzle is concerned, the particular senses Fred associates with 'Twain' and 'Clemens' are neither here nor there: They can be swapped without consequence. Contrast this with a position more like Peacocke's, on which the content of Fred's Twain-concept would not only determine its reference but would also be what explained the rationality of certain basic inferences in which that concept was involved. (I do not mean to say that Peacocke himself would subscribe to this particular view.) In retrospect, it is easy enough to see how the description theory of names satisfied this condition—consider, for example, the allegedly *a priori* status of 'Hesperus is sometimes visible in the evening'—but, of course, the description theory has its own problems, and few Fregeans would endorse it nowadays. Still, one lesson of the present paper is that, if the notion of sense is to be defended, then the reference-fixing and rationalist elements are essential: Only with them in place will there be work for the notion of sense to do that cannot be done by the sorts of formal relations to which

[81] That Evans discusses language at all in *Varieties* sometimes seems to me an accident of intellectual culture.

[82] As said earlier, the possession condition for disjunction is that one accept instances of these rules as 'primitively compelling' in virtue of their form.

I have argued we must appeal in solving Frege's puzzle.

That Frege's own understanding of the notion of sense is shaped by his rationalism has been a theme in much of Tyler Burge's recent work on Frege (Burge, 2005b,a). And, as we have just seen, this connection is honored in the work of some self-described Fregeans. At least some of us have believed, however, or at least have hoped, that the notion of sense itself could be loosed from these epistemological moorings and reconstructed in purely cognitive terms. If I am right, then we were wrong.[83]

References

Aydede, M. (2000a). 'Computation and intentional psychology', *Dialogue* 39: 365–79.

—— (2000b). 'On the type/token relation of mental representations', *Facta Philosophica: International Journal for Contemporary Philosophy* 2: 23–49.

Aydede, M. and Robbins, P. (2001). 'Are Frege cases exceptions to intentional generalizations?', *Canadian Journal of Philosophy* 31: 1–22.

Beck, J. (2008). *The Structure of Thought*. PhD thesis, Harvard University.

Bilgrami, A. (1998). 'Why holism is harmless and necessary', *Philosophical Perspectives* 12: 105–26.

Block, N. (1986). 'Advertisement for a semantics for psychology', *Midwest Studies in Philosophy* 10: 615–678.

Boghossian, P. (2003). 'Blind reasoning', *Proceedings of the Aristotelian Society* sup. vol. 77: 225–48.

Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge MA, Harvard University Press.

Braun, D. (1998). 'Understanding belief reports', *Philosophical Review* 107: 555–95.

[83] Thanks to Jake Beck, Chris Hill, Jim Pryor, and Michael Rescorla for discussion of these issues that significantly influenced my thinking about them. Talks based upon this paper were presented at the University of Pittsburgh, in December 2006; at the University of Cincinnati and the Univeristy of Chicago, in May 2007; at Rutgers University in November 2007; and at Wake Forest University, in February 2008. Thanks to the participants at all of these events for their comments and suggestions, but especially to Ruth Chang, Cian Dorr, David Finkelstein, Chris Gauker, Stavroula Glezakos, Anil Gupta, and Ernie Lepore. Thanks also to an anonymous referee, whose comments did much to improve the paper.

—— (2000). 'Russellianism and psychological generalizations', *Nous* 34: 203–236.

—— (2001). 'Russellianism and explanation', *Philosophical Perspectives* 15: 253–89.

Burge, T. (2005a). 'Frege on apriority', in Burge 2005c, 356–87.

—— (2005b). 'Frege on knowing the foundation', in Burge 2005c, 317–55.

—— (2005c). *Truth, Thought, Reason: Essays on Frege*. New York, Oxford University Press.

Cappelen, H. and Lepore, E. (1997). 'On an alleged connection between indirect speech and the theory of meaning', *Mind and Language* 12: 278–296.

Carroll, L. (1895). 'What the tortoise said to Achilles', *Mind* 4: 278–80.

Chomsky, N. (2000). 'Naturalism and dualism in the study of language and mind', in *New Horizons in the Study of Language and Mind*. New York, Cambridge University Press, 75–105.

Davidson, D. (1984a). *Inquiries Into Truth and Interpretation*. Oxford, Clarendon Press.

—— (1984b). 'Radical interpretation', in Davidson 1984a, 125–139.

—— (1984c). 'Thought and talk', in Davidson 1984a, 155–70.

Davies, M. (1992). 'Aunty's own argument for the language of thought', in J. Ezquerro and J. M. Larrazabal (eds.), *Cognition, Semantics, and Philosophy*. Boston, Kluwer Academic Publishers.

—— (1998). 'Language thought, and the language of thought (Aunty's own argument revisited)', in P. Carruthers and J. Boucher (eds.), *Language and Thought: Interdisciplinary Themes*. Cambridge, Cambridge University Press, 226–47.

Dennett, D. (1971). 'Intentional systems', *Journal of Philosophy* 68: 87–106.

Dummett, M. (1978). 'The social character of meaning', in *Truth and Other Enigmas*. London, Duckworth, 420–430.

—— (1981). *Frege: Philosophy of Language*, 2d edition. Cambridge MA, Harvard University Press.

Evans, G. (1982). *The Varieties of Reference*, McDowell, J., ed. Oxford, Clarendon Press.

—— (1985). 'Understanding demonstratives', in *Collected Papers*. Oxford, Clarendon Press, 291–321.

Fiengo, R. and May, R. (1994). *Indices and Identity*. Cambridge MA, MIT Press.

Fine, K. (2007). *Semantic Relationism*. Blackwell.

—— (2010). 'Comments on Scott Soames' 'Coordination problems'', *Philosophy and Phenomenological Resarchy* 81: 475–84.

Fodor, J. (1975). *The Language of Thought*. Cambridge MA, Harvard University Press.

—— (1982). 'Cognitive science and the twin-earth problem', *Notre Dame Journal of Formal Logic* 23: 98–118.

—— (1987). *Psychosemantics*. Cambridge MA, MIT Press.

—— (1994). *The Elm and the Expert*. Cambridge MA, MIT Press.

—— (1998). *Concepts: Where Cognitive Science Went Wrong*. New York, Oxford University Press.

—— (2003). *Hume Variations*. Oxford University Press.

Frege, G. (1984a). *Collected Papers on Mathematics, Logic, and Philosophy*, McGuiness, B., ed. Oxford, Basil Blackwell.

—— (1984b). 'Function and concept', tr. by P. Geach, in Frege 1984a, 137–56. Also in Frege 1997, 130–48.

—— (1984c). 'On sense and meaning', tr. by M. Black, in Frege 1984a, 157–77. Also in Frege 1997, 151–71.

—— (1997). *The Frege Reader*, Beaney, M., ed. Oxford, Blackwell.

Gabbay, D. (1998). *Fibring Logics*. Oxford, Oxford University Press.

Grimm, R. and Merrill, D., eds. (1988). *Contents of Thought*. Tuscon, University of Arizona Press.

Harman, G. (1988). *Change in View: Principles of Reasoning*. Cambridge MA, MIT Press.

Heck, R. G. (1995). 'The sense of communication', *Mind* 104: 79–106.

—— (2000). 'Non-conceptual content and the "space of reasons"', *Philosophical Review* 109: 483–523.

—— (2002). 'Do demonstratives have senses?', *Philosophers' Imprint*, 2. `http://www.philosophersimprint.org/002002/`.

—— (2003). 'Frege on identity and identity-statements: A reply to Thau and Caplan', *Canadian Journal of Philosophy* 33: 83–102.

—— (2005). 'Reason and language', in C. MacDonald and G. MacDonald (eds.), *McDowell and His Critics*. Oxford, Blackwells, 22–45.

—— (2007a). 'Are there different kinds of content?', in J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in Philosophy of Mind*. Oxford, Blackwells, 117–38.

—— (2007b). 'Meaning and truth-conditions', in D. Greimann and G. Siegwart (eds.), *Truth and Speech Acts: Studies in the Philosophy of Language*. New York, Routledge, 349–76.

Heck, R. G. and May, R. (2006). 'Frege's contribution to philosophy of language', in E. Lepore and B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*. Oxford, Oxford University Press, 3–39.

—— (2010). 'The composition of thoughts', *Noûs* 45: 126–66.

Higginbotham, J. (1985). 'On semantics', *Linguistic Inquiry* 16: 547–593.

—— (1989). 'Knowledge of reference', in A. George (ed.), *Reflections on Chomsky*. Oxford, Basil Blackwell, 153–174.

—— (1992). 'Truth and understanding', *Philosophical Studies* 65: 3–16.

Kaplan, D. (1978). 'Dthat', in P. Cole (ed.), *Pragmatics*. New York, Academic Publishers, 221–243.

Kripke, S. (1976). 'A puzzle about belief', in A. Margalit (ed.), *Meaning and Use*. Dordrecht, Reidel, 239–83.

Larson, R. and Ludlow, P. (1993). 'Interpreted logical forms', *Synthese* 95: 305–355.

Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. Cambridge MA, MIT Press.

Lewis, D. (1986). *On the Plurality of Worlds*. Cambridge MA, Blackwell.

Loar, B. (1988a). 'A new kind of content', in Grimm and Merrill 1988, 122–39.

—— (1988b). 'Social content and psychological content', in Grimm and Merrill 1988, 99–110.

Mates, B. (1952). 'Synonymity', in L. Linsky (ed.), *Semantics and the Philosophy of Language*. Champaign IL, University of Illinois Press, 111–36.

May, R. (2006). 'The invariance of sense', *Journal of Philosophy* 103: 111–144.

McDowell, J. (1998). 'On the sense and reference of a proper name', in *Meaning, Knowledge, and Reality*. Cambridge MA, Harvard University Press.

McGinn, C. (1982). 'The structure of content', in A. Woodfield (ed.), *Thought and Object: Essays on Intentionality*. Oxford, Clarendon Press, 207–58.

Monk, J. D. (1986). 'The contributions of Alfred Tarski to algebraic logic', *Journal of Symbolic Logic* 51: 899–906.

Peacocke, C. (1992). *A Study of Concepts*. Cambridge MA, MIT Press.

—— (1993). 'Externalist explanation', *Proceedings of the Aristotelian Society* 93: 203–30.

—— (1998). 'Implicit conceptions, understanding, and rationality', *Philosophical Issues* 9: 45–88.

—— (2004). *The Realm of Reason*. Oxford, Oxford University Press.

Perry, J. (1993). 'Frege on demonstratives', in *The Problem of the Essential Indexical, and Other Essays*. New York, Oxford University Press, 3–32.

Putnam, H. (1975). 'Brains and behavior', in *Mind, Language, and Reality*. Cambridge, Cambridge University Press, 325–41.

Quine, W. V. O. (1953). 'Two dogmas of empiricism', in *From a Logical Point of View*. Cambridge MA, Harvard University Press, 20–46.

Richard, M. (1990). *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. New York, Cambridge University Press.

Rives, B. (2009). 'Concept cartesianism, concept pragmatism, and Frege cases', *Philosophical Studies* 144: 211–38.

Salmon, N. (1986). *Frege's Puzzle*. Cambridge MA, MIT Press.

Schneider, S. (2005). 'Direct reference, psychological explanation, and Frege cases', *Mind and Langauge* 20: 423–47.

Segal, G. (1997). 'Content and computation: Chasing the arrows', *Mind and Language* 12: 490–501.

—— (2000a). 'Frege's puzzle as some problems in science', *Rivista di Linguistica* 8: 375–388.

—— (2000b). *A Slim Book About Narrow Content*. Cambridge MA, MIT Press.

Segal, G. and Sober, E. (1991). 'The causal efficacy of content', *Philosophical Studies* 63: 1–30.

Siegel, S. (2006). 'Subject and object in the contents of visual experience', *Philosophical Review* 115: 355–88.

Soames, S. (1987). 'Direct reference, propositional attitudes and semantic content', *Philosophical Topics* 15: 47–87.

—— (2010). 'Coordination problems', *Philosophy and Phenomenological Research* 81: 464–74.

Stalnaker, R. (1984). *Inquiry*. Cambridge MA, MIT Press.

—— (1999a). *Context and Content: Essays on Intentionality in Speech and Thought*. New York, Oxford University Press.

—— (1999b). 'The problem of logical omniscience, I', in Stalnaker 1999a, 241–54.

—— (1999c). 'The problem of logical omniscience, II', in Stalnaker 1999a, 255–73.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge MA, MIT Press.

Taschek, W. (1992). 'Frege's puzzle, sense, and information content', *Mind* 101: 767–791.

—— (1995). 'Belief, substitution, and logical structure', *Noûs* 29: 71–95.

—— (1998). 'On ascribing beliefs: Content in context', *Journal of Philosophy* 95: 323–353.

Thau, M. (2002). *Consciousness and Cognition*. Oxford, Oxford University Press.

Thau, M. and Caplan, B. (2001). 'What's puzzling Gottlob Frege?', *Canadian Journal of Philosophy* 31: 159–200.

Wakefield, J. (2002). 'Broad versus narrow content in the explanation of action: Fodor on Frege cases', *Philosophical Psychology* 15: 119–33.

Williamson, T. (2003). 'Understanding and inference', *Proceedings of the Aristotelian Society* 77: 249–93.