

# Intuition and the Substitution Argument

*Richard G. Heck, Jr.*

**Department of Philosophy, Brown University**

Superman has a bad allergy to kryptonite. And since Superman is Clark Kent, it follows that Clark Kent is allergic to kryptonite. Which, indeed, he is. If you were to take some kryptonite into the offices of the *Daily Planet*, Clark would soon be feeling quite poorly. Of course, not many people *know* that Clark is allergic to kryptonite. In particular, Superman's nemesis Lex Luthor does not know that Clark is allergic to kryptonite, which is why he does not sneak into the newsroom and hide some kryptonite in Clark's desk. But of course, Luthor knows perfectly well that Superman is allergic to kryptonite, and he frequently tries to expose him to it.

That, at least, is the sort of thing one would usually hear people say. Familiarly, however, taking such remarks at face value generates problems. The most immediate of these is that it requires us to restrict the logical principle known as "substitution of identicals":

$$A(t), t = u \vdash A(u)$$

Otherwise, we should be licensed to reason as follows:

(S) Lex Luthor knows that Superman is allergic to kryptonite.

(=) Superman is Clark Kent

(C) Lex Luthor knows that Clark Kent is allergic to kryptonite.

But that seems invalid, since the premises seem to be true and the conclusion seems to be false. Moreover, it would appear that the contribution a proper name makes to the truth-condition of a sentence in which it occurs is not always determined entirely by its reference: "Superman" and "Clark Kent" have the same reference, and yet (S) and (C) have different truth-values and so different truth-conditions.

Another consequence concerns the semantics of certain sorts of constructions. How extensive the failure of substitution might be is not entirely clear, but the sort of example we are discussing involves names that occur within the complement of a verb attributing a ‘propositional attitude’: “knows”, in this case, but also “believes”, “desires”, “fears”, and so forth. And, from this point of view, what the ‘substitution argument’ most immediately threatens is the view now known as ‘Russellianism’, which we might formulate as the conjunction of the following two principles:

- (i) Propositional attitude verbs, such as “knows” and “believes”, serve to relate a thinker to the semantic value of the ‘that-clause’ occurring as complement of the verb. That is: “*N* knows that *S*” is true if, and only if, the reference of “*N*” bears the knowing relation to the semantic value of “that *S*”.
- (ii) The semantic value of a phrase of the form “that *S*” is the Russellian proposition expressed by the embedded sentence *S*.

Exactly what a ‘Russellian proposition’ might be is not crucial for our purposes. Indeed, all that really matters is that which Russellian proposition is expressed by a sentence depends, so far as any names that might occur in that sentence are concerned, only upon their reference.<sup>1</sup> But, if so, then “that Superman is allergic to kryptonite” has the same semantic value as “that Clark Kent is allergic to kryptonite”. And then it follows from (i) that (S) and (C) must have the same truth-value. But that, again, seems wrong, so Russellianism must be false.

In fact, the problem is not limited to that-clauses but concerns complements of other forms. For example, Luthor wants to destroy Superman, but he does not want to destroy Clark Kent; Lois expects Superman to save the children, but she does not expect Clark to save the children; and so forth.<sup>2</sup> These sorts of examples involve non-finite clauses. In the

<sup>1</sup> So the view that “that *S*” denotes the set of metaphysically possible worlds in which *S* is true is equally threatened by the substitution argument.

<sup>2</sup> A quite different case is: Lois persuaded Superman to save the children. Here, the structure is: Lois persuaded Superman [PRO to save the children]. “Superman” is thus the object of the verb “persuaded”, and it controls the null subject PRO in the complement. It is not clear to me whether, if Lois persuaded Superman to save the children, she also persuaded Clark to save the children. The issue here is not about the direct object. If Lois persuaded Superman that Boston is lovely, then she persuaded Clark that Boston is lovely. The issue concerns what PRO inherits, in such a context, from its controller and so is similar to the issues presented by a case like: Lois told Clark [resp., Superman] that Jimmy said he was a twit. A particularly interesting version of

former case, the subject of the clause is the null element PRO, which is ‘controlled’ by “Luthor”. The grammatical structure is generally supposed to be: Luthor wants [PRO to destroy Superman], which has roughly the meaning: Luthor wants Luthor to destroy Superman. For simplicity, however, I’ll stick in what follows to more familiar sorts of cases.

I should emphasize that what matters here is that utterances of (C) *can* be false even when corresponding utterances of (S) would be true. It is no part of my view, and I see no reason it should be part of anyone’s view, that there are not contexts in which utterances of, e.g., “Lois knows that Clark Kent can fly” might be true, even though Lois does not know that Clark is Superman. I suspect that the truth of a sentence of the form “*N* knows that *S*” does not require *N* to know *the very proposition* that is expressed by the embedded sentence *S*, and similarly for other propositional attitudes.<sup>3</sup> It is enough if *N* knows (believes, etc) some proposition that is appropriately related to that expressed by *S*, and what counts as an appropriately close relation depends upon context.<sup>4</sup> That relation could well be loose enough, in a particular context, that an utterance of “Lois knows that Clark Kent can fly” would be *de re* with respect to “Clark Kent” and so mean, roughly: Lois knows, of Clark Kent, that he can fly. But I shall ignore those complications, as well, in what follows.

To return to our main thread, then, the entire weight of the substitution argument rests upon the claim that (S) can be true even if (C) is false. What is the basis for that claim? The usual view seems to be that it is based upon a ‘linguistic intuition’, or a ‘truth-conditional intuition’. It was in deference to this sort of idea that I was careful, above, to claim only that it *seems* as if (S) is true and (C) is false. A different, but related, thought would be that the fact that ordinary speakers frequently say such things as

(F) Luthor knows that Superman is allergic to kryptonite, but not that Clark is.

---

this sort of case, noted by Robert May, is: Superman said that he could fly, but Clark didn’t. This exhibits a familiar strict–sloppy ambiguity, and one might think the sloppy reading could be true.

<sup>3</sup> With Bach (2000), I thus reject what he calls the ‘Specification Assumption’, though for quite different reasons.

<sup>4</sup> Russellians, note, might hold a similar view, but they cannot allow that utterances of (S) and (C) even *could* have different truth-values in a single context, since context will affect the interpretations of the common elements of (S) and (C) the same way, but cannot affect what reference “Superman” and “Clark Kent” have.

is evidence, of some kind, that the (S) and (C) have different truth-conditions. But that seems to come, in the end, to the same thing, except that now we seem to be relying not just upon our own ‘intuitions’, but upon those of ordinary speakers as well.

If, however, we have nothing but an ‘intuition’ that (S) can be true when (C) is false, then there is an obvious strategy for responding to the substitution argument: One need only explain these ‘intuitions’ away. Or, if we are simply appealing to facts about what ordinary speakers tend to say, then we need only explain why speakers might utter (F), despite the fact that it is not true. And, of course, there is a familiar way of doing that: Claim that, although what is *said* by an utterance of (F) cannot be true, what is *communicated* by an utterance of (F) can be true; that is, utterances of (F) tend, in Grice’s sense, to implicate things that are true, and these implicatures are what are of interest to ordinary speakers.<sup>5</sup> This ‘pragmatic reply’ to the substitution argument has been defended by several philosophers, including Nathan Salmon (1986), Jennifer Saul (1998), and Scott Soames (1987).

I am going to argue, however, that the best version of the substitution argument is one in which ‘intuitions’ do not play any role that would allow the argument to be answered by explaining those ‘intuitions’ away. More precisely, the basis for the claim that (S) can be true when (C) is false is not that we have an ‘intuition’ that they can be. There is, rather, an *argument* to be given that (S) can be true even when (C) is false. One cannot respond to that argument by ‘explaining away’ whatever ‘intuitions’ there might be that happen to comport with its conclusion. No pragmatic explanation of why ordinary speakers might utter (F) even if it is false answers the argument, either, since the argument offers reason to think that utterances of (F) can be true that does not rest solely upon what ordinary speakers tend to say.<sup>6</sup>

<sup>5</sup> There are other options, too. One might hold that, for some reason, ordinary speakers are just wrong about whether (S) can be true when (C) is false. Of course, one wants to know what explains such widespread error, but perhaps there is a story to be told. Indeed, it is Saul’s central purpose in her book *Simple Sentences, Substitution, and Intuitions* (Saul, 2007) to tell just such a story. Our focus here will be on the pragmatic reply, though, for reasons that will emerge below. See note 34.

I should emphasize that nothing I will say here undermines Saul’s claim that, in principle, there are ways to explain why people regularly utter sentences that are not true besides claiming that they thereby communicate things that are true. That is an important observation.

<sup>6</sup> I have made this sort of complaint before (Heck, 1995, p. 80, fn. 4), but am only now developing it properly.

I will approach this issue by first discussing, in Section 1, an influential objection to the substitution argument due to Saul (1997). This will serve to highlight the role played by 'intuition' in discussions of the substitution argument, since Saul's objection rests crucially upon the claim that 'intuitions' *do* play a central role in that argument. In Section 2, I will distinguish the questions at issue in the substitution argument, which concern the attribution of beliefs, from corresponding questions about the nature of belief itself, and then rehearse an argument, closely related to the substitution argument, that, e.g., the belief that Superman is allergic to kryptonite is distinct from the belief that Clark is. In Section 3, I will digress briefly to consider whether a weaker claim than the one made in Section 2 would suffice for what follows, though we will continue to operate with the stronger claim. In Section 4, I will then discuss an argument of Frege's that attempts to parlay the argument of Section 2 into one for the conclusion that (S) can be true even when (C) is false. As we shall see, Frege's version of this argument depends upon a stronger assumption than is warranted by the arguments of Section 2, and it is, in any event, insufficiently developed. In Section 5, then, I will develop Frege's argument myself and defend it against a series of objections. I'll summarize the results in Section 6.

## 1 Saul's Puzzle

As we have seen, the substitution argument rests upon the claim that substitution of co-referential expressions, inside the clauses that specify the contents of attitudes, can fail to preserve truth-value. In her much discussed paper "Substitution and Simple Sentences", however, Saul presents a series of examples that seem, *prima facie*, to show that similar failures of substitution occur even outside attitude attributions.

It frequently happens in the Superman stories, for example, that, Clark, while walking down the street and playing the mild-mannered reporter, spies a dangerous situation in which he needs to intervene. So he dashes into a phone booth, changes into his superhero outfit, and emerges to save the day. It seems natural, in such a situation, to say something like:

(P) Clark Kent went into the phone booth, and Superman came out.

But it does not seem at all natural, in the same circumstances, to say:

(P?) Superman went into the phone booth, and Superman came out.

Similarly, given Superman's heroism, it is easy to imagine that the local restaurant owners are more than happy to sit the Man of Steel at their best table. But reporters gets no such treatment. So it seems as if it could be reasonable to say:

(T) Superman gets better tables than Clark Kent.

But, of course, it would never be reasonable to say:

(T?) Superman gets better tables than Superman.

And, as Saul (1997, p. 103) notes, the problem doesn't just arise with superheroes, and it isn't even confined to people. It is easy to imagine circumstances in which someone might say:

(L) I never made it to Leningrad, but I visited St. Petersburg last week.

But no-one would ever say:

(L?) I never made it to Leningrad, but I visited Leningrad last week.

Additional such cases are easily generated.

Saul's point in discussing these examples is *dialectical*. There are two options for responding to them. The first option is to argue that we can, and should, take the appearances at face value: Utterances of (T) can be true even though utterances of (T?) cannot be. The second option is to insist that (T) cannot be true and then to explain why it is, nonetheless, sometimes a reasonable thing to say—or, at least, why people sometimes say it, even people who know that Superman is Clark Kent.

There have been several philosophers who have wanted to go the first route, including Graeme Forbes (1997; 1999), Joseph Moore (1999), David Pitt (2001), and Stefano Predelli (1999; 2004). As one might imagine, there are all sorts of problems confronting this approach, some of which Saul (1997, pp. 103–7) discusses in her original paper. The central difficulty is that it is hard to implement this option without threatening the truth of identity claims such as "Superman is Clark Kent". For example, you might take "Superman" (sometimes) to refer to time-slices in which a certain person is wearing blue tights and a cape, and "Clark Kent" (sometimes) to refer to time-slices in which that same person is wearing a suit and nerdy glasses. But then it is not clear why no utterance of "Superman is Clark Kent" can be false, since the former time-slices are distinct from the latter. But if such identity claims aren't

true, then we don't have a puzzle. And surely they are true. Are Clark and Superman both supposed to get a vote in the Metropolis elections?

In any event, the first option strikes me as desperate. There are good reasons to insist that (T) and (T?) cannot differ in truth-value, namely, that substitution of co-referential expressions, at least outside attitude contexts, is truth-preserving (Braun and Saul, 2002, p. 2). The destruction the first option visits upon the simple semantics for proper names to which we are otherwise attracted can only be warranted if there is no better option.<sup>7</sup>

And one might well have thought that the second option was obviously the right choice. Now, however, is when Saul springs the trap she has been setting for her opponent. If we go that way, then it seems that we have no reason to resist the pragmatic reply to the substitution argument itself. As Saul puts it:

The proponent of this [second] response. . . now has a choice. She must decide whether or not to accept *a perfectly parallel account of our intuitions about attitude reports*, [namely:] substitution of co-referential names in attitude reports preserves truth conditions but may result in the generation of new, and misleading, pragmatic implicatures. It is these implicatures which result in our (mistaken) tendency to say that [(S)] may be true while [(C)] is false. . . . The main argument against [this] theory has been that it requires the violation of our *intuitions* about substitution. But the current approach to [(T) and (T?)] requires what is apparently *a perfectly parallel violation of intuitions*, accompanied by *a perfectly parallel appeal to pragmatics*. (Saul, 1997, pp. 106–7, my emphases)

We have, Saul says, an 'intuition' that (T) can be true, even though (T?) cannot be. The second option is to explain this 'intuition' pragmatically: Although (T) can never be true, utterances of (T) can nonetheless communicate things that are true, by the usual sorts of mechanisms.<sup>8</sup> But

<sup>7</sup> The idea that "Superman" has an odd semantics might have some plausibility, but the idea that "Clark Kent" does has none. Yet the truth of the identity statement "Superman is Clark" would still be threatened, even if only one of these names had an odd semantics.

<sup>8</sup> Alex Barber (2000) develops a version of this proposal, which both Braun and Saul (2002, pp. 9-11) and Saul (2007, ch. 3) discuss in some detail. Their criticisms rest upon views about implicature—in particular, about the role played by speakers' intentions—

then, the thought is, we should just respond to speakers' tendency to utter such things as

( $\bar{C}$ ) Luthor does not believe that Clark is allergic to kryptonite.

in the same way: ( $\bar{C}$ ) cannot be true when (S) is true, but speakers utter it because it (sometimes) implicates things that are true and that they wish to communicate. As Saul (1997, p. 107) concludes the passage quoted above: "The advocate of [the second option] owes us a reason for supposing that one set of intuitions deserves to be taken so much more seriously than the other". Or, as she would prefer: The advocate of the second option should stop claiming that the substitution argument poses some special problem for Russellian views about attitude ascription. It's a more general problem, and it needs a more general solution.

I (doubly) highlighted the three occurrences of the word "intuition" in the long passage quoted above to emphasize the central role Saul supposes 'intuitions' to play in the substitution argument.<sup>9</sup> In particular, Saul is clearly supposing that the substitution argument is based upon an 'intuition' that (S) can be true when (C) is false—an 'intuition' that might then be 'explained away'. So far as I am able to determine, however, Saul nowhere says what she means by an 'intuition'. For our purposes, though, we do not need to know. What matters here is just Saul's insistence that the cases of (S) and (C), on the one hand, and of (T) and (T?), on the other, are "perfectly parallel"—a phrase she uses three times in the long passage quoted above. In particular, it is essential to Saul's argument that the *basis on which it is claimed* that (C) can be false even when (S) is true should be the same as the *basis on which it is claimed* that (T) can be true, although (T?) cannot be. Only if that is so will our giving a pragmatic explanation of why people sometimes say things like (T) have any tendency to commit us to offering a 'parallel' explanation of why people sometimes utter (S) when they would also be prepared to utter ( $\bar{C}$ ).

It is essential here to distinguish the claim that (S) can be true when (C) is false, which we might call a *linguistic judgement*, and whose subject-matter is a *linguistic phenomenon*, from the *linguistic data* that

---

that I have argued elsewhere are false (Heck, 2005, §1). That said, I would not endorse Barber's particular view. He too seems to me to assume too much about how implicature works.

<sup>9</sup> In the paper Braun and Saul wrote together, the word "intuition" occurs, by my count, almost a hundred times, just in the main body of the paper.



might support such a judgement.<sup>10</sup> Such data could, in principle, include all sorts of things. But, for our purposes, the most important linguistic data are claims about the circumstances in which various sentences could appropriately be uttered, claims that might be based upon observations about the circumstances in which those sentences are in fact uttered. Data are never theory-neutral, of course, but data should be as neutral between competing theories as possible. That is why I characterized the data in terms of when a sentence can be uttered *appropriately*, not *truly*. Notions like truth, or the contrast between what is said and what is meant, enter only at the level of linguistic phenomena.

With this distinction in mind, then, let us ask on what basis it is supposed to be claimed that (T) can be true.

I hope that sounds like an odd question. It certainly does to me. I'm not sure we have any reason at all to claim that (T) can be true. The only relevant datum is that people, even people who know that Superman is Clark Kent,<sup>11</sup> sometimes say such things as (T). But that fact, by itself, does not seem to me to provide much support at all for the claim that (T) can be true. No doubt, if (T) is not capable of truth, then one will want an explanation of why people utter such falsehoods—even, again, people who know that Superman is Clark Kent. If there were no such explanation to be had, then that might be reason to think that (T) can be true after all. But the *only* burden one assumes if one rejects the claim that utterances of (T) can be true is to explain why people do in fact utter (T).

If the case of (S) and (C) is to be “perfectly parallel”, then, the basis for the claim that (S) can be true when (C) is false must simply be that people, even people who know that Superman is Clark Kent, sometimes utter (S) when they would also be prepared to utter ( $\bar{C}$ ). But that is not the only basis for the claim that (S) can be true when (C) is false. On the contrary, there is an *argument* to be given for that claim, one I shall spend the rest of this paper elaborating. So, if one wants to reject that claim, it is not sufficient to explain why people—even people who know that Superman is Clark Kent—sometimes utter (S) when they would also be prepared to utter ( $\bar{C}$ ). You have to answer the argument. Facts about what people are inclined to say in different circumstances will prove to be relevant to that argument, but, as we shall see, no pragmatic

<sup>10</sup> I borrow some of this terminology from Peter Ludlow (2011, §3.1). See his discussion for more detail on this distinction.

<sup>11</sup> Of course people who do not know that Superman is Clark might say things like (T). But we need no special explanation of that fact.

explanation of those facts could help one to answer it. And, if that is right, then the case of (T) and (T?) and the case of (S) and (C) are *not* parallel in the way Saul thinks they are.

We can now see why the notion of a ‘truth-conditional intuition’ is so slippery. In so far as a ‘truth-conditional intuition’ is supposed to concern *truth-conditions*, its subject-matter is an alleged linguistic phenomenon. But when people speak of ‘intuitions’, they generally mean something that would count as data and, more precisely, something that is readily available and that “constrast[s with] the judgements we produce through deliberative reasoning” (Nagel, 2012, p. 498).<sup>12</sup> So use of the term ‘truth-conditional intuition’ tends to elide the distinction between linguistic phenomena and linguistic data and so to suggest that we have a kind of immediate insight into such matters as whether a sentence could *truly*—and not just *appropriately*—be uttered in certain circumstances. So far as I can see, however, there is no reason to think either that we do have such ‘immediate insight’ or that the actual practice of natural language semantics presumes that we do.<sup>13</sup>

It is in that sense, then, that I will be arguing that ‘intuition’ need play no role in the substitution argument: The basis for the crucial claim that (S) can be true when (C) is false is not that, if one said otherwise, that would conflict with some ‘intuition’ that we all share. So, even if it is true that “the main argument against [Russellianism] *has been* that it

<sup>12</sup> In her book—which she advertises as showing “traditional approaches to truth-conditional intuitions to be inadequate” (Saul, 2007, p. ix)—Saul spends a fair bit of time discussing the question what sorts of ‘intuitions’ people really have, suggesting that work of the sort now familiar from so-called ‘experimental philosophy’ might throw light on the matter (Saul, 2007, §6.10, esp. §6.10.3). One hypothesis, then, is that ‘intuitions’, as Saul understands them, are much the same sort of thing that experimental philosophers have called by that name, and that is the sort of conception Nagel is elaborating in the remarks quoted in the text.

<sup>13</sup> One might think the contrary view was expressed in passages like this one, from Jason Stanley and Zoltan Gendler Szabó (2000, p. 240):

The obvious disadvantage [of pragmatic approaches to quantifier domain restriction] is that one has to abandon ordinary intuitions concerning the truth or falsity of most sentences containing quantifiers. This is worrisome because accounting for our ordinary judgements about the truth-conditions of various sentences is the central aim of semantics.

But, while Stanley and Szabó are not as sensitive to the distinction between linguistic phenomena and linguistic data as they might be, I doubt that their considered view would conflict with the points being elaborated in the text. In any event, they explicitly set the alleged conflict with ‘intuition’ aside and give a very different sort of argument against pragmatic approaches: the much discussed ‘binding argument’.

requires the violation of our *intuitions* about substitution” (Saul, 1997, p. 107, my emphasis), there is a better argument to be had. And I’m not sure it is true. Frege’s own argument, I’ll suggest below, does not seem to have been of that form.

## 2 Substitution and Belief

I now begin to reconstruct the substitution argument.

We need first to distinguish between questions about *belief* and questions about *belief-attribution*.<sup>14</sup> When I speak about belief, I am talking about certain content-bearing mental states that are causally implicated in the production of action. Questions about the nature of such states are empirical questions: ones that are, ultimately, in the domain of cognitive psychology and the brain sciences. When I speak about belief-attribution, by contrast, I am talking about certain expressions of natural language, e.g. the verb “believe” as it is used in such constructions as “*N* believes that *S*”. Questions about belief-attribution are questions about the way such expressions work. These too are empirical questions, but ones that fall within the domain of linguistics and, more precisely, of natural language semantics.

The distinction is particularly important here because Russellianism of the sort that Saul is defending against the substitution argument is a view about *belief-attribution*, not a view about belief itself. The Russellian view about belief-attribution is, as was said earlier, that “believes” expresses a binary relation between a thinker and the Russellian proposition that is the semantic value of the complement clause. It follows, then, that (S) and (C) cannot differ in truth-value. But Russellians such as Braun, Salmon, Saul, and Soames do *not* hold the corresponding view about belief itself, which I shall call the Naïve View: that beliefs, those very mental states, are binary relations between thinkers and Russellian propositions. On the contrary, Russellians about belief-attribution tend to be more Fregean about belief itself, for reasons we shall explore shortly.

I want to emphasize, before we continue, however, that it is *not* patently absurd to think that our ordinary use of the verb “believe” might not precisely track the facts about beliefs themselves, e.g., that ordinary speech might obscure distinctions that matter to psychology. Or, perhaps

---

<sup>14</sup> Barbara Partee (1988, p. 47) nicely emphasizes the dangers of conflating these two sorts of questions.

better: It is not patently absurd to think that the everyday concept of belief might differ, in significant ways, from whatever similar concept might be required for the purposes of genuine scientific explanation. Indeed, it would be surprising if there were no such divergences. Though scientific concepts often emerge from corresponding concepts deployed by the folk, the ordinary concept almost always has to undergo significant refinement before it is useful for the unordinary purposes of empirical science. To mention just one familiar, and quite obvious, respect in which the folk concept of belief would have to be refined: Ordinary usage does not make any sharp distinction between implicit and explicit belief, any more than ordinary usage clearly distinguishes mass from weight, heat from temperature, or average from instantaneous velocity.

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 “believe” simply because the verb “believe” expresses the relation of belief. But this claim rests upon assumptions about how the verb “believe” works that are not in evidence. In particular, it assumes that “*N* believes that *S*” will be true only if that utterance of *S* has the very same content as some *belief* of *N*’s, where the italicized occurrence of the word “belief” expresses the *scientific* concept of belief. But that is a very strong claim, and I see no reason to regard it as true. Still, one might wonder just how great a divergence between the facts about belief and the semantics of “believe” it is possible to tolerate. We shall return to that issue in Section 5.

Now, as I said above, Russellians about belief-attribution do not usually endorse the Naïve View that beliefs themselves are binary relations between thinkers and Russellian propositions. Indeed, almost no-one endorses the Naïve View. The reason is that there is a familiar argument, one that is very much like the substitution argument, that seems to most of us to refute the Naïve View. The argument goes roughly as follows. Imagine that Lois is sitting at her desk at the offices of the *Daily Planet* and talking on the phone to Clark, who is about to cover a school board meeting. A few minutes later, Jimmy Olson pokes his head through the door and tells her, “Someone just said they saw Superman at the school board meeting!”<sup>15</sup> Excitedly, Lois hangs up on Clark, jumps up from her

<sup>15</sup> This is only one way of telling the story, of course, and language is not essential to it. Lois might have learned that Superman was at the school board meeting in all kinds of ways, and the rest of the argument would then proceed without change. For example, Lois might have been watching a television broadcast of the meeting on Metropolis’s public access channel and seen Superman there herself. Or she might have arrived at

desk, and rushes out the door. Why?

Lois, as always, desperately wants to interview Superman. So, one might have thought, the reason Lois rushes out the door is because she has reason to believe that Superman is at the school board meeting. More precisely, when Jimmy said what he did, Lois came to have a belief about Superman's location, one she did not previously have. And that new belief—that Superman is at the school board meeting—is what explains Lois's subsequent behavior. But Lois believed all along that Clark was at the school board meeting. And so, according to the Naïve View, she already believed, even before Jimmy said what he did, that Superman was at the meeting. That makes it a natural question why Lois did not rush out the door the moment Clark told her where he was, and it is not clear what answer the Naïve View has to offer.

It is important to understand that this argument does *not* rest upon an 'intuition' that, before Jimmy said what he did, Lois believed that Clark was at the meeting, but not that Superman was. It is no doubt true that ordinary speakers are inclined to say such things as:

Lois rushed out because, when Jimmy said what he did, Lois came to believe that Superman was at the meeting, which she had not previously believed, though she had previously believed that Clark was at the meeting.

But the objection that is being brought against the Naïve View is *not* that it makes claims about what Lois believes that are incompatible with what ordinary people would be inclined to say (i.e., that it has to reject this 'intuition', or 'explain it away'). The objection is that the Naïve View prohibits us from offering any cognitive explanation of why Lois behaves as she does. One might try just accepting this conclusion, as Jerry Fodor (1994, ch. 1) once did. Since psychological laws are not exceptionless, Fodor's idea was, perhaps Frege cases are simply among the exceptions. Agents' behavior in such 'Frege cases' is thus *cognitively* inexplicable, though of course their behavior will be explicable in other terms, e.g., neurologically. As Gabriel Segal (1997) has argued, however, Fodor's view has serious problems of detail, and Fodor has since abandoned it.

But if we are not to deny that Lois's behavior is cognitively explicable, the question stands: What cognitive change does Lois undergo, when

---

this conclusion by an inference to the best explanation, based upon other facts she knew. She might even have come to believe the Superman was at the meeting for no particular reason at all. What matters is only that she does come to believe it.

Jimmy says what he does, that might explain her change of behavior? And, if we reject the Naïve View, we can easily answer that question, thus: Lois came to believe that Superman was at the meeting. But that answer, again, is unavailable to the Naïve View. Of course, there are all kinds of other answers a defender of the Naïve View might suggest. But I have argued elsewhere that none of those other answers will do (Heck, 2012, §2). And, as I have said, Russellians about belief-*attribution* generally accept this sort of argument against the Naïve View, so there is no need to defend it in detail here. One of the replies that is available to the Naïve View will be particularly relevant later, however, so we should discuss it briefly now.

That reply, which I call the ‘Braun Variation’, is that Lois’s behavior should be explained not by her coming to believe that Superman is at the meeting but, say, by her coming to believe that *the guy who can save Metropolis* is at the meeting. That is a belief that the Naïve Theorist can perfectly well allow that Lois did not previously have, so long as she insists that it is not a singular belief but a descriptive one.

One worry about this reply is that it threatens to make the Naïve View look a good deal more like Russell’s actual epistemology than I would have supposed its proponents intended: If this reply is adequate, then the crucial cognitive difference in Frege cases must always lie in agents’ *descriptive* beliefs, since, according to the Naïve View, there can never be a difference, in such cases, in the agents’ *singular* beliefs. But if so—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 (Segal, 2000). Are we to say that it *would* have been appropriate to explain Lois’s rushing out in terms of her then coming to believe that Superman was at the meeting if, by some chance, she had not previously known that Clark was there? But that, since she did know that Clark was there, we cannot explain her rushing out in that way? That simply does not seem plausible.<sup>16</sup> What if Lois had been watching TV and had seen Clark in disguise, not recognizing him? Then she would have had a demonstrative belief that *that guy* is at the

---

<sup>16</sup> It would be a mistake to think that such plausibility judgements signal an appeal to ‘intuition’ in the sense in which Saul uses the term. By contrast, George Bealer (1996; 1998) thinks of ‘intuition’ as a special faculty of rational insight and argues, in effect, that our endorsement of the premises of our arguments often rests upon intuition in that sense. As it happens, I would reject this view for much the same reasons Williamson (2008, esp. ch. 6) and Cappelen (2012, esp. §6.1) do. But the present point is simply that ‘intuition’ in this sense is not at all what Saul has in mind.

meeting, and again her behavior would have to be explained in terms of her descriptive beliefs.

There is, however, a more serious objection, namely, that there are other Frege cases with which this reply cannot deal. Indeed, it generates a Frege case with which it cannot deal. We need only ask *why* Lois comes to believe that the guy who can save Metropolis is at the meeting when Jimmy says what he does if she did not so believe before. The Naïve Theorist claims that Lois did previously know that Superman was at the meeting. But she has known for years that Superman is the guy who can save Metropolis. Why, then, did she not make a simple inference and conclude that the guy who can save Metropolis was at the meeting? Or, better: Why was she *in no position* to make such an inference? Why would it have been irrational for her to do so, if she had? The problem for the Naïve View, once again, is not just that it prohibits us from answering these questions in the way we can answer them if we reject it. The problem is that the Naïve View prevents us from giving *any* cognitive explanation of Lois's change of mind.<sup>17</sup>

If we are not simply to decline to give any cognitive explanation of how agents behave in Frege cases, then, we must allow that, when Jimmy said what he did, Lois acquired a new belief, one she did not previously have: It is Lois's having this new belief that explains why she jumped up from her desk and rushed to the school board meeting. So we have to replace the Naïve View with some other account of the metaphysics of belief. There are several options. One is to retain the idea that belief is a binary relation between a thinker and a proposition, but to make propositions more fine-grained. That is the Fregean view. Another is to retain the Russellian conception of propositions, but to take belief to be a ternary relation between a thinker, a proposition, and a 'way of apprehending' that proposition. That is the view we find in Braun (1998, pp. 564–5), Salmon (1986, pp. 105–14), and Soames (1995, pp. 522–4). Alternatively, one might take a belief to be a relation between a thinker, a mental representation, and the proposition that is the content of that representation. The proposition in question might yet be a Russellian proposition, and one would then explain speakers' behavior in Frege cases in terms of the fact that they are storing information about a single object in different formats. Different forms of this view have been defended by Fodor (1994), by me (Heck, 2012), and by Susan Schneider

<sup>17</sup> For a more thorough discussion, see §2.1 of "Solving Frege's Puzzle". It is particularly important to note that the argument does not depend upon any kind of closure principle.

(2005; 2011), and it has obvious affinities with views that invoke so-called ‘mental files’, such as that of François Recanati (2013). It is also the sort of view on which Saul (2007, §6.1) relies.

It does not matter for our purposes which of these views one prefers. What matters here is simply that this argument for the fine-grainedness of belief in no way depends upon claims about what people are or would be inclined to say in various real or imagined circumstances. It’s an argument about how beliefs have to be individuated if they are to play an appropriate role in psychological explanation, or the causation of behavior, or whatever you think the right notion to use here is.

So, to summarize: What has been argued so far is that there are two different mental states into which Lois can get herself, states that play different roles in her psychology. One is the belief that is responsible for Lois’s getting up from her desk and rushing to the school board meeting. The other is the belief that does not, in these circumstances, cause her to rush to the meeting but that might, in other circumstances, be responsible for her going to the meeting if, say, she needed to find another reporter to help her cover a breaking story. We might call the former belief *the belief that Superman is at the meeting* and the latter *the belief that Clark is at the meeting*. In the present context, however, those labels would be tendentious. So we shall call the former *the S-belief* and the latter *the C-belief*. And, just to be clear, I have *not* argued that these two beliefs have different contents. The only thing I have argued is that they are different mental states, since they play different roles in Lois’s psychology.

### 3 A Brief Digression

The argument just given depends, as it was stated, upon the assumption that beliefs play a role in the causation, or explanation, of rational behavior. But of course, not everyone agrees that beliefs play *any* role in human psychology. Eliminativists, for example, do not think there are any beliefs at all, so they would certainly reject the argument just given for the distinctness of the S-belief and the C-belief. By itself, of course, that does not show that there is anything wrong with the argument. It just means that among its premises is the claim that the overall picture of the mind that posits such mental states as beliefs, desires, and the like—states that interact causally in ways that mirror relations between their contents—is not just a ‘folk’ theory no more deserving of epistemic



respect than is ‘folk’ physics.

In the remainder of this section, I will consider whether, for the purposes of this paper, we might do with a weaker assumption than the one just mentioned. This question seems to me to be important, but nothing in the rest of the paper depends upon the answer, so those who wish to do so can skip ahead to Section 4.

As it happens, I actually do think that there are genuine cases in which someone believes that  $a$  is  $F$  but not that  $b$  is  $F$ , even though  $a$  is  $b$ . That is because I do not myself believe that eliminativism is true, and I accept the arguments sketched in Section 2. But, even if eliminativism is true, surely it isn’t *necessarily* true. So, even if eliminativism is true, the arguments sketched above (if otherwise correct) still show that the S-belief and the C-belief are distinct, in the sense that someone *could* have the S-belief without having the C-belief. Of course, that will only be so in ‘possible worlds’ in which things are broadly as ‘folk’ psychology would have them be, i.e., in which people’s behavior is indeed to be explained in terms of their beliefs, desires, and so forth. Since the ordinary use of such words as “believes” obviously presupposes that there are such things as beliefs in terms of which people’s behavior can be explained, that will be enough for what follows.

A different, but perhaps related, response would note that it is one thing to suppose that there really are beliefs, and it is another thing to suppose that ordinary people have a *concept* of belief or, more generally, a battery of folk psychological notions in terms of which they understand and explain their own behavior and that of others. It is, of course, an empirical question what our everyday ‘theory of mind’ is actually like.<sup>18</sup> But, however much may still remain to be discovered, it seems to me that we have excellent evidence already for the following claim:

(TM) Normal human beings who have achieved a certain level of cognitive development distinguish, within their ‘theory of mind’, between, e.g., Lois’s having the S-belief and her having the C-belief, since they regard Lois as liable to behave in different ways depending upon which of these beliefs she has, and this is true even if they themselves know that Superman is Clark Kent.

<sup>18</sup> The still classic work on these issues is *The Child’s Theory of Mind*, by Henry Wellman (1990), though of course much has happened since. I know of no comparable, more recent survey written by a psychologist, but there is a nice ‘popular’ discussion by Alison Gopnik (2009, esp. ch. 2) in her book *The Philosophical Baby*, and Susan Carey (2009, esp. ch. 5) touches on these topics in *The Origin of Concepts*, as well. Alvin Goldman (2012) surveys relevant work from a more philosophical perspective.

Indeed, the sorts of arguments sketched above are readily adapted to support (TM). What ultimately lies behind those arguments is the observation that the Naïve View “force[s] us to give very different explanations of what would otherwise seem to be very similar occurrences” (Heck, 2012, p. 135). But that will be true whether the explanations in question are being offered by cognitive scientists or by ordinary people. Suppose, as before, that Lois knew that Clark was at the meeting, but this time Jimmy says, “Lex Luthor is at the meeting!” Out Lois goes again. Ordinary people have no problem understanding why. Their theory of mind provides for a difference between Lois’s believing that Clark is at the meeting and Lois’s believing that Luthor is at the meeting, and they use that distinction to make sense of Lois’s behavior more or less as follows:

When Jimmy said what he did, Lois came to believe that Lex Luthor was at the meeting, which she hadn’t previously believed, etc, etc.

And, as a matter of empirical fact, people treat Frege cases no differently. No one feels as if they have to go searching for some totally different sort of explanation in the original case involving Superman.<sup>19</sup>

So there is good reason to suppose that our theory of mind provides us with the conceptual resources needed to distinguish between Lois’s having the S-belief and her having the C-belief, even if we ourselves know that Superman is Clark. That is, there is good reason to think that (TM) is true. And that, again, will be enough for what follows. The substitution argument is supposed to show that (S) and (C) have different truth-conditions. It won’t matter for that argument whether, as a matter of fact, anyone has any beliefs. What will matter is that ordinary speakers think they do.<sup>20</sup>

To be sure, (TM) is an empirical claim, and so, in principle, closer examination of how human beings make sense of Frege cases could show that our theory of mind is itself Russellian. I doubt it. But the primary question here is not whether this or that version of the substitution

---

<sup>19</sup> Suppose we tell the story about Clark and Superman to someone who does not know that Clark is Superman and then tell them that he is. Would one expect him to think it necessary to revise his previous understanding of why Lois rushed out of the office? No, we’d expect him to say something like, “I guess Lois didn’t know that Clark was Superman”.

<sup>20</sup> Of course, eliminativists would deny that ordinary speakers think anything, and they would be unlikely to care much about empirical work in cognitive science on ‘theory of mind’. But that is a somewhat different issue. Eliminativists aren’t likely to care much about the semantics of belief-ascription, either.

argument should be regarded as successful, but whether the best version of that argument must appeal to ‘intuition’ in the way Saul supposes it does. So, in particular, one will want to know whether it is possible to defend (TM) without appealing to ‘intuitions’ that might be ‘explained away’. But the sort of argument sketched in favor of (TM) does not, it seems to me, appeal to ‘intuition’ in anything like the way the ‘argument’ that (T) can be true does. How ordinary people actually respond to Frege cases is indeed relevant to the argument for (TM), but as a guide to the conceptual resources that our ‘theory of mind’ makes available to us.

#### 4 Frege on Belief and Belief-Attribution

I said earlier that I would be arguing that the basis on which it is claimed, as part of the substitution argument, that

(LC) Lois believes that Clark is at the school board meeting.

can be true even if

(LS) Lois believes that Superman is at the school board meeting.

is false is not at all the same as the basis on which one might claim that (T) can be true even though (T?) cannot be. Now, that (LC) can be true even if (LS) is false is a claim about belief-attribution, not about belief itself. So we need to do some more work if we are to extend the conclusions of Section 2, which concern belief, to conclusions about the language we use to talk about belief.

There is one very quick way to complete the argument that (LC) can be true even if (LS) is false. Suppose we were to assume that (LC) expresses the proposition that Lois has the C-belief (i.e., the belief that Clark is at the meeting), and that (LS) expresses the proposition that Lois has the S-belief (i.e., the belief that Superman is at the meeting). Then it would follow immediately that (LC) can be true even when (LS) is false. That is because we already know that it is possible for Lois to have the C-belief when she does not have the S-belief, as indeed she did before Jimmy said what he did. In those circumstances, that is to say, (LC) is true and (LS) is false, and so we would be done.

I strongly suspect that Frege, and many others who have written on this topic, have quietly made exactly this sort of assumption. Indeed, Frege’s own argument in “On Sense and Reference” seems to have a structure very similar to that of the one just sketched. Frege does not

actually give much of an argument that, e.g., (LC) can be true when (LS) is false. He simply writes:

In indirect speech one talks about the sense, e.g., of another person's remarks. It is quite clear that in this way of speaking words do not have their customary meaning but designate what is usually their sense. (Frege, 1984c, op. 28)

One might think that Frege is simply taking it to be obvious that (LC) can be true when (LS) is false, and that the words “It is quite clear. . .” report an ‘intuition’ to that effect. But such a suggestion would ignore the fact that this remark occurs four pages into “On Sense and Reference” and that Frege has, to that point, given *plenty* of reason to suppose that there is a distinction between sense and reference. What sits at the foundation of his argument for that distinction, or so I would argue (Heck, 2002, §1; 2003, §2), is the claim that the *thought* that Hesperus is a planet is different from the *thought* that Phosphorous is a planet, since someone “might regard one as true and the other as false” (Frege, 1984b, op. 14).<sup>21</sup> Frege offers no real argument that these thoughts are distinct. Rather, it just seems obvious to Frege that someone might believe the thought that Hesperus is a planet while not believing the thought that Phosphorous is a planet. Of course, one might reasonably demand an argument for this claim—which is, note, one about belief, not about attribution—and the last section can be thought of as an attempt to provide one on Frege's behalf.<sup>22</sup>

For our purposes, however, the important point is how Frege argues *from* this starting point *to* the conclusion that (LC) can be true when (LS) is false. Assuming that there are distinct thoughts about Venus to the effect that it is a planet, Frege proceeds to argue that the sentences “Hesperus is a planet” and “Phosphorous is a planet” express different of these thoughts: the H-thought and the P-thought, respectively. With this conclusion in hand, Frege then remarks:<sup>23</sup>

<sup>21</sup> Note that what Frege calls a “thought” is the *content* of an episode of thinking, not the episode itself. So thoughts play the same sort of role for Frege that propositions do for others.

<sup>22</sup> Though, as noted above, the argument falls short of what Frege needs, since it does not imply that the S-belief and the C-belief have distinct contents. There is discussion of why one might be tempted to draw that conclusion, and how it might be resisted, in “Solving Frege's Puzzle” (Heck, 2012, §§II-III).

<sup>23</sup> I've adjusted the translation here to use “reference” rather than “meaning” as the translation of “Bedeutung”.

If words are used in the ordinary way, what one intends to speak of is their reference. It can also happen, however, that one wishes to talk about the words themselves or their sense. (Frege, 1984c, op. 28)

For example, if someone has made an assertion, one might want to report what they have said, and, given what Frege has already argued, that will require one to identify the thought expressed by their words. The implicit question is how it is possible to do this: How, in particular, might one report that someone has asserted the H-thought as opposed to the P-thought? Frege's answer is that we can do so via "indirect speech": That is the mechanism language provides for us to "talk[] about the sense. . . of another person's remarks". Frege's argument, that is to say, rests upon the claim that, e.g., "Jimmy said that Hesperus is a planet" expresses the proposition that Jimmy asserted the H-thought, whereas "Jimmy said that Phosphorous is a planet" expresses the proposition that Jimmy asserted the P-thought. But, if so, then it is indeed "quite clear that in this way of speaking words do not have their customary meaning but designate what is usually their sense" (Frege, 1984c, op. 28).

So, to sum up, if we assume, with Frege, that (LC) expresses the proposition that Lois has the C-belief, and that (LS) expresses the proposition that Lois has the S-belief, then we can easily extend the argument given in the last section to reach the conclusion that (LC) can be true when (LS) is false. To be sure, Frege does not give much of an argument for the additional assumption just stated, and one might well want such an argument. But Frege does not say nothing. He claims that we sometimes have an interest in discussing the thought someone has expressed; he asks what resources language provides for this purpose; and he then finds those resources more or less where we might have expected to find them: in indirect speech.

It is at least not obvious that this argument of Frege's rests, any more than the argument of Section 2, on anyone's 'intuitions'. But one might well wonder whether an appeal to 'intuition' would emerge if we were to develop Frege's argument about what is expressed by indirect speech. Moreover, Frege's argument depends upon claims about the distinctness of *thoughts*—that is, of contents—whereas I noted at the beginning of this section that the argument given in Section 2 does not yield any such conclusion, but only one about the distinctness of certain *mental states*. So we need to develop Frege's argument if only to determine whether it can be adapted to this weaker premise.

The more important point, however, is that, if Frege's argument is directed against Russellian views of belief-ascription, then it seems almost to beg the question. It is absolutely central to such views to deny that the semantics of the verb "believe" (or "say") tracks the metaphysics of belief (or assertion) in the way Frege supposes it does. Contemporary Russellians agree that the S-belief is distinct from the C-belief, but they insist nonetheless that (LS) and (LC) must be true or false together. So they must, and happily do, deny that (LS) expresses that Lois has the S-belief specifically. Their view is that (LS) expresses that Lois believes the Russellian proposition expressed by "Superman is at the meeting" under some 'guise' or other<sup>24</sup> and that (LC) expresses that Lois believes the Russellian proposition expressed by "Clark is at the meeting" under some 'guise' or other. Since the two quoted sentences express the same Russellian proposition, any belief whose content fits the former specification will also fit the latter specification. That is why (LS) and (LC) must be true or false together. If one is going to argue against this position, then, one needs to say quite a bit more than Frege does in defense of the claim that (LS) expresses the proposition that Lois has the S-belief specifically and not, as Russellians would have it, that she has either the S-belief, or the C-belief, or any one of a number of other such beliefs.

Now, I insisted earlier that it is not patently absurd to suppose that belief and belief-attribution come apart in the way that Russellians think they do. But it is a different question whether they *do in fact* come apart in that way. This question is especially pressing given how similar the argument for the fine-grainedness of belief rehearsed in Section 2 looks to be to the substitution argument. It may well be consistent to accept the former but reject the latter, as Russellians do, but it is not at all obvious that it is coherent to do so.<sup>25</sup> And, or so I am about to argue, it is not: The sort of argument Frege gives for the claim that a clause

---

<sup>24</sup> This is stronger than is actually required, since Russellians about ascription could hold Fregean views about belief itself. In that case, we would say that the 'Russellian part' of the content of the belief was a certain Russellian proposition. In fact, however, I don't think there is much difference between Frege's view and views that make use of 'guises' (Heck, 2012, pp. 146–9).

<sup>25</sup> A brief argument against Russellianism based on this sort of consideration was presented in the original version of "Solving Frege's Puzzle", but it had to be removed due to considerations of length. What remained was simply a footnote suggesting that the arguments against the Naïve View could be transposed to an argument against Russellianism (Heck, 2012, p. 137, fn. 9). Those who are interested can find the longer version on my website.

in indirect speech refers to its sense can be transferred to the present setting.

## 5 From Belief to Belief Attribution

Speakers use language to serve many different sorts of interests. And, since we humans are social creatures, it is unsurprising that one of the things in which we are sometimes interested is how our fellow creatures can be expected to behave in different sorts of circumstances. For example, Lois's friends and acquaintances are liable to be interested in how she can be expected to act, so that is something they may sometimes wish to talk about.

Now, we know from Section 2 that the S-belief and the C-belief are distinct, in the sense that it is possible for Lois to have one of these beliefs without having the other. We also know that Lois may behave differently depending upon which of these beliefs she has.<sup>26</sup> It would therefore be unsurprising if ordinary speakers sometimes had reason to distinguish between these two circumstances: Lois's having the S-belief, and Lois's having the C-belief. The obvious question to ask, then, is a variant of the question we saw Frege wanting to ask: How is it possible, in English (and other natural languages), to express that Lois has the S-belief as opposed to the C-belief?

It is of course a presupposition of this question that natural languages such as English are not so impoverished that this sort of difference cannot be expressed in them. I have no argument to offer in defense of this claim, other than to observe that languages grow and change to serve the expressive and communicative needs of their users, and so that it would be extremely surprising if there simply was, in ordinary English, no way to express such a significant difference as the one we are discussing. It is not a response to note that the difference could be expressed if we were to introduce new vocabulary that allowed us to talk about modes of presentation or mental files (or what have you), so that the difference between Lois's having the C-belief and her having the S-belief could be expressed by well-informed philosophers and cognitive scientists (Braun, 1998, fn. 23).<sup>27</sup> For one thing, in our present state of knowledge, even

<sup>26</sup> At least in worlds in which beliefs play the kind of role in the causation of behavior that our ordinary ways of speaking presuppose they do. See Section 3.

<sup>27</sup> The same goes, of course, for the language we are using here: the "S-belief" and the "C-belief". (Remember that we are only using such language because the obvious language to use would be tendentious in the present setting.)

the best-informed philosophers and cognitive scientists have only a very rough idea how to talk about modes of presentation and mental files. For another, that it might one day be possible to introduce technical language allowing us to express the difference between Lois's having the S-belief and her having the C-belief does nothing to counter-balance the implausibility of the suggestion that, until then, it will be impossible for ordinary human speakers to do so. If the distinction were subtle, one that emerged only from extensive philosophical or empirical investigation, then that would be different. But that is very much not the situation.

Indeed, one good reason to think that the distinction between the S-belief and the C-belief is expressible in English is the fact that human speakers seem to make this very distinction all the time. Faced with a Frege case, we are not simply dumbstruck when asked to explain the agent's behavior. Rather, we say things like:

Lois rushed out because, when Jimmy said what he did, Lois came to believe that Superman was at the meeting, which is something she had not previously believed, though she had previously believed that Clark was at the meeting.

And someone who says such a thing certainly looks as if she is trying to distinguish between Lois's having the S-belief and her having the C-belief: She is trying to say that Lois already had the C-belief before Jimmy said what he did; that she came to have the S-belief when he spoke; and that it was the S-belief, rather than the C-belief, that caused Lois to rush out the door. So, I boldly suggest, the way people actually speak in such circumstances gives us reason to suppose that:<sup>28</sup>

- The sentence (LS), "Lois believes that Superman is at the school board meeting", expresses the proposition that Lois has the S-belief.

<sup>28</sup> Note that the real question here is how it is possible to register the *difference* between Lois's having the S-belief and her having the C-belief. What we want is thus a sentence for whose truth Lois's having the S-belief is sufficient, but for whose truth her having the C-belief is not sufficient. I claim that (LS) is such a sentence. We do *not* require a sentence that is true just in case Lois has the S-belief. As David Sosa emphasized to me, if the contents of beliefs are extremely fine-grained, there may be no such sentence. It is, however, much easier to speak as I do in the text, so I shall continue to speak that way, to simplify the exposition.

Remember, as well (see page 3), that the claim here is not that utterances of (LS) always *must* express that Lois has the S-belief as opposed to the C-belief, but only that utterances of (LS) *can* express that Lois has the S-belief rather than the C-belief. These weaker claims are not available to the Russellian any more than the stronger claim in the text is.



- The sentence (LC), “Lois believes that Clark Kent is at the school board meeting”, expresses the proposition that Lois has the C-belief.

And, as we saw in Section 4, if that is true, then it follows that (LC) can be true even when (LS) is false, since, prior to Jimmy’s speaking, Lois does have the C-belief but does not have the S-belief.

Now, this argument clearly does appeal to facts about what people are inclined to say in certain sorts of circumstances. But it is not being argued that, if some view were to reject the claim that (LS) expresses the proposition that Lois has the S-belief, that would be a mark against it, since that would not ‘accord with intuition’. Facts about what people say are being invoked, rather, in the course of answering the question how a certain difference—between Lois’s having the S-belief and her having the C-belief—might be expressed in English. The methodology is to consider a case in which ordinary speakers might have an interest in expressing this difference and then to look at what sorts of things they are inclined to say. That cannot but provide some evidence—it is not, of course, conclusive evidence—about how the difference in which we are interested might be expressed.

To see the contrast, consider Gettier cases. Here is what a parallel treatment of them would be like. First, we would need already to have an independent argument that there was a difference between justified true belief and knowledge. Second, we would need reason to believe that ordinary speakers sometimes have an interest in distinguishing someone’s knowing something from their believing it both truly and with justification. That would make it plausible that this difference should be expressible in English, so it would then be reasonable to ask how it might be expressed. Third, we would need to establish that Gettier cases are ones in which the subject has a justified true belief but does not know. Then, but only then, could we appeal to what people are inclined to say about Gettier cases to support the claim that the word “know”, in English, expresses the relation of knowledge, rather than that of justified true belief.<sup>29</sup>

<sup>29</sup> It is, no doubt, relevant to such an argument what people *are* inclined to say about Gettier cases, and that is obviously an empirical question. It is equally relevant what people are in fact inclined to say about Frege cases, and that is also an empirical question. But there are plenty of real-life Frege cases, and I take it that we know what sorts of thing people tend to say about them. To mention several, in his book *Black Like Me*, John Howard Griffin recounts how differently he was treated when he traveled across the Deep South in 1959 after darkening his skin so that he would present as African-American. Many of the people he encountered knew him as a white man but

Now, to be sure, not everyone thinks that the philosophical significance of Gettier cases depends upon our ‘intuitive judgments’ about them (Williamson, 2008, ch. 6; Cappelen, 2012, pp. 193–4), and I tend to agree. But how Gettier cases *should* be understood is not relevant here. What is relevant is that people who *do* think that our ‘intuitive judgments’ are what matter cannot understand Gettier cases in the way I have just described. Why, after all, are our ‘intuitive judgments’ about these cases supposed to be so much as relevant? Because they are supposed to provide evidence about when someone is correctly described, in English, as ‘knowing’ something. Gettier (1963) was arguing that *there is* a distinction between knowledge and justified true belief, not exploring how that distinction, independently drawn, might be expressed in English. By contrast, I am not appealing to what people are inclined to say to argue that we should distinguish the S-belief from the C-belief. The argument for that claim was given in Section 2.<sup>30</sup>

Nor is the fact that speakers utter (LC) and (LS) in different circumstances, *by itself*, being treated as evidence that these sentences express different propositions, let alone that what they express are, respectively, that Lois has the C-belief and that she has the S-belief. We only have reason to ask how that difference can be expressed in English because we have, again, independent reason to believe *both* that there is such a difference *and* that this difference is of interest to ordinary speakers. And it is only because it is so difficult to see how else that difference *could* be expressed in English—to see what other sentence, if not (LS), might express that Lois has the S-belief specifically—that what people are inclined to say gives us reason to suppose that that is what (LS) expresses. And it is, for the same reason, no response to insist that (LC) and (LS) must have the same truth-value and then to try to explain, as Braun (1998) and Saul (2007) do, why ordinary speakers might wrongly think otherwise. That can at most show, e.g., that (LS) does not express that Lois has the S-belief specifically. It does not answer the question what sentence *does* express that she has the S-belief.

---

failed to recognize him as a black man. There is no sense disputing that people do in fact say such things as “Sterling Walker didn’t realize that he was talking to John Howard Griffin”.

<sup>30</sup> If we construe the argument along the lines indicated at the end of Section 3, then we would be appealing to what speakers are inclined to say about Frege cases in order to argue that our ‘theory of mind’ provides for a distinction between the S-belief and the C-belief. Such a use of what speakers are inclined to say is independent, however, of whether utterances of (LS) *say* or just *mean* that Lois has the S-belief. And, in principle, there could be all sorts of other evidence for the availability of that distinction, as well.

Similarly, then, the mere fact that ordinary speakers are inclined to utter (T) in certain circumstances, but are never inclined to utter (T?), gives us no reason, *by itself*, to believe that (T) and (T?) express different propositions. I suppose that, if pressed, I might be willing to grant that it is *some* evidence that they do.<sup>31</sup> But, as Grice (1989) long ago pointed out, it is at best defeasible evidence, and I myself would regard it as quite weak. In any event, in the present context, the crucial question, once again, is how it is *possible* to express a certain difference. If there were some other way, besides uttering (LS), in which ordinary speakers might attribute the S-belief to Lois, then that would undermine the argument I have given for the claim that (LS) expresses that Lois has the S-belief. So what we need to do in the case of (T) and (T?) is to show, or at least to make it plausible, that there are other ways, besides uttering (T), in which one might express that circumstances are such as might prompt an utterance of (T). Which there does seem to be, e.g.: Clark (i.e., Superman), when he is dressed in such a way that people recognize him as Superman, gets better tables than when he is not so dressed.<sup>32</sup> On the other hand, as I have said, I have no idea what sentence other than (LS) might express that Lois has the S-belief specifically.

Still, one might object that I have no right to assume that ordinary speakers need to be able to *express* that Lois has the S-belief.<sup>33</sup> One might grant that ordinary speakers might take an interest in the difference between Lois's having the C-belief and her having the S-belief, and so that they might wish to communicate that Lois has the S-belief specifically. But they can do that, the objection would continue, so long as utterances of (LS) implicate that Lois has the S-belief specifically. The pragmatic strategy is thus still available.<sup>34</sup>

The first point to note is that, if this reply is intended as complete in its own right, and not to be augmented by any suggestion about how the

<sup>31</sup> Presumably, it raises the probability that (T) and (T?) express different propositions at least slightly.

<sup>32</sup> Is that exactly what is meant by utterances of (T)? I doubt there is anything that is exactly meant by such utterances. That is quite generally true when one communicates by implicature, as I've emphasized elsewhere (Heck, 2005, §1).

<sup>33</sup> Or, more precisely, that there are sentences that *express* the difference between Lois's having the S-belief and her having the C-belief. See note 29.

<sup>34</sup> Note that only the pragmatic strategy helps here. In particular, the view that ordinary speakers are for some reason wrong about whether (LC) can be true when (LS) is false—a view defended, in different ways, by Braun (1998), Braun and Saul (2002), and Saul (2007)—does not help to answer the objection that Russellianism makes certain explanatory facts inexpressible.

difference in which we are interested might be *expressed*, then it has the consequence that utterances of (LS) sometimes implicate a proposition that it is utterly impossible actually to express in English. I know of no antecedent for such a claim, and it strikes me as extremely implausible, on general grounds.<sup>35</sup> And I am not alone. Braun (1998, pp. 567–8) makes exactly this sort of objection to the pragmatic account.<sup>36</sup> But let us set that complaint aside for the moment and explore a different reply to the objection.

Not only is the S-belief is different from the C-belief, but how Lois behaves may differ depending upon which of these beliefs she has. To put it differently, there are situations in which Lois's behaving in a certain way is due to her having the S-belief, rather than the C-belief, or in which her not behaving in a certain way is due to her not having the S-belief, though she does have the C-belief. These sorts of facts are ones in which ordinary speakers might well take an interest. Indeed, the reason ordinary speakers sometimes wish to distinguish between Lois's having the S-belief and her having the C-belief is precisely because of the difference these beliefs make to her behavior, which is to say that ordinary speakers have reason to be interested in the question what difference it will make to Lois's behavior now that she has the S-belief, or what difference it would make if she did have it. These, then, are things that ordinary speakers might well wish to discuss. For example, they might wish to say that it was due to Lois's having the S-belief (not the C-belief) that she rushed out of the office; that her not rushing out earlier was due to her not having the S-belief, although she did then have the C-belief; and so forth. It is thus a natural question how such things can be expressed in English.

If that sounds familiar, then of course it should. The sort of argument given above is simply being transposed now to a setting in which we are interested in expressing, not just the difference between Lois's mental states, but how her being in those different states affects her behavior. And, as before, the question how it is possible to express, in English, that some action of Lois's was due to her having the S-belief, or to

---

<sup>35</sup> This is a form of Searle's "Principle of Expressibility" (Searle, 1969, p. 20). It is questioned by Camp (2006, pp. 14ff), on the ground that metaphor is an exception. That case, however, seems very different from this one.

<sup>36</sup> There are obvious similarities between the objection I am offering and an earlier one due to François Recanati. His argument turns, however, on the requirement that we be able consciously to separate what is said and what is meant (Recanati, 1993, p. 245). Here, the thought is just that what we can mean we should also be able to say.

her not having the C-belief, of course presupposes that such things are expressible in English.<sup>37</sup> But, as before, it seems to me exceedingly unlikely that it should not be, just on general grounds, and all the more so since ordinary speakers often do seem to note the different effects that such beliefs have. That is, to repeat, ordinary speakers frequently say such things as:

Lois rushed out *because*, when Jimmy said what he did, Lois came to believe that Superman was at the meeting.

Someone who says such a thing certainly looks as if she is at least trying to say that Lois's acting as she did was due to her having the S-belief rather than the C-belief. So, if we consider these sentences:

(B1S) Lois rushed out because she believed that Superman was at the school board meeting.

(B2S) Lois remained at her desk because she did not believe that Superman was at the school board meeting.

then, I suggest, the way people actually speak in such circumstances gives us reason to suppose that (B1S) expresses the proposition that Lois rushed out because she had the S-belief and that (B2S) expresses the proposition that Lois remained at her desk because she did not have the S-belief.

Russellians, however, hold that (B2S) expresses the very same proposition as:<sup>38</sup>

(B2C) Lois remained at her desk because she did not believe that Clark Kent was at the school board meeting.

But all parties to the debate agree that (B2C) is false, since (i) Russellians and non-Russellians alike agree that "Lois believes that Clark is at the meeting" was true all along and (ii) a statement of the form "*A* because *B*" cannot be true unless both *A* and *B* are true, i.e., explanatory statements of this sort are factive. So, according to Russellians, (B2S) must also be false. But then (B2S) cannot express that Lois remained at her desk because she lacked the S-belief, since then it would have to be true, since

<sup>37</sup> Here again, the critical question is how it is possible to express the *difference* between Lois's acting as she did because she had the S-belief and her acting as she did because she had the C-belief. See note 29.

<sup>38</sup> Braun (2001, p. 271) makes this point explicitly. But then he makes things too easy for himself by not considering the crucial case of non-belief.

that is indeed why she remained at her desk. But what we wanted to know was how the causal relation between Lois's beliefs and her actions could be expressed in English. If it is not (B2S) that expresses that Lois remained at her desk because she lacked the S-belief, then what sentence of English would?<sup>39</sup>

It is no response to this argument to insist that (B2S) and (B2C) are both false and then to try to explain, as Braun (2001, p. 273) does, why someone might think otherwise. The question that is at issue is how certain causal (or explanatory) relations can be expressed in English—relations we have independent reason to believe obtain and are of interest to ordinary speakers. It is only because it is so difficult to see how else those relations *could* be expressed in English—to see what other sentence, if not (B2S), might express that Lois remained at her desk because she lacked the S-belief—that what people are inclined to say gives us reason to suppose that that is what (B2S) expresses. The fact that people are inclined to utter such things as (B2S) in certain sorts of circumstances certainly is relevant to the argument, then. But it is not, by itself, being treated as evidence that (B2S) and (B2C) have different truth-conditions, so explaining why someone might wrongly think they could have different truth-values is beside the point.

Now, again, one could accept this conclusion and insist that (B2S) does not express but only implicates that Lois rushed out because she had the S-belief. But unless one can identify some other way in which that proposition might be expressed, one will then have committed oneself to the view that there are propositions that, as Braun puts it, “ordinary speakers routinely entertain and believe” but that they “do not have a conventional way of expressing”. Indeed, one might wonder not only why we have never “develop[ed] some ordinary expression or construction that can be used for this purpose” (Braun, 1998, p. 568) but how ordinary speakers ever come to be able to entertain such propositions in the first place.

That said, however, one might simply deny that such explanatory relations can be expressed in English, at least in some cases. Braun does just that:

... [I]f Russellianism is correct, then there may be some *unusually recalcitrant cases* for which *no ordinary attitude ascription* can provide contrastive explanatory information

---

<sup>39</sup> Similar considerations apply to counterfactuals, such as: If Lois had known that Superman was at the meeting, she would have gone.

about the agent's... behavior. (Braun, 2001, p. 278, my emphasis)

The cases Braun has in mind are ones involving what John Perry (1993) called "essential indexicals", such as that of the amnesiac lost in the Stanford library. To my mind, however, these sorts of cases are not at all unusual but simply illustrate important features of such phenomena as self-conscious knowledge, features that are always present even if they are not always obvious. So, if such cases are "unusually recalcitrant" from Braun's point of view, perhaps that says more about Braun's point of view than it does about the cases themselves. Still, perhaps Braun would agree with Herman Cappelen and Josh Dever (2014) that there is, on closer examination, no such thing as essential indexicality, in which case there would be nothing left unexplained.

I disagree, but we do not have to decide that issue, because there is a more immediate worry about Braun's position. I am happy to concede that it is not utterly implausible that English should lack the resources needed to express suitable explanations of people's behavior in some cases. But the plausibility of this move is inversely proportional both to the number and to the importance of those cases. That is why it is essential that Braun should restrict his bullet-biting to "some unusually recalcitrant cases". The question thus arises what Braun means to do about all the other, perfectly ordinary cases, such as the ones we have been discussing. The answer is that he proposes that we should explain Lois's getting up from her desk and rushing out of the office not, as we have been supposing, in terms of her having the S-belief but, rather, in terms of her having some *descriptive* belief, say, that *the guy who can save Metropolis* is at the school board meeting (Braun, 2001, pp. 277–8), which we'll call the M-belief. And if that is right, then of course English does have the resources to express a suitable explanation of Lois's behavior: Lois rushed out because she came to believe that the guy who can save Metropolis was at the school board meeting.

This view is closely related to the 'Braun variation' of the Naïve View, for which it was the inspiration. But Braun's actual view is not, in fact, a variant of the Naïve View, so it is not vulnerable to the argument against the Braun variation that was presented in Section 2. That argument involved posing the question why Lois did not have the M-belief prior to Jimmy's saying what he did: Since Lois already believed, before Jimmy spoke, that Clark was at the meeting, the Naïve View is committed to her already believing as well that Superman was at the meeting. Since she

has always known that Superman is the guy who can save Metropolis, it is then a mystery why she didn't already believe that the guy who can save Metropolis was at the meeting. But Braun rejects the Naïve View.<sup>40</sup> He accepts that the S-belief is distinct from the C-belief. So Braun can happily answer the questions just posed to the Naïve View by noting that, before Jimmy said what he did, Lois had neither the S-belief nor the M-belief. She only came to have the S-belief when Jimmy spoke. And then, having the S-belief, she was able to infer the M-belief from it together with her prior knowledge that Superman is the guy who can save Metropolis.

Now, however, in the immortal words of Yogi Berra, it's *déjà vu* all over again.<sup>41</sup> These sorts of relations between Lois's beliefs seem to be something in which ordinary speakers might well take an interest. They have, after all, an interest in how her coming to have the S-belief might affect her behavior, and one way in which it will do so is indirectly: by affecting what else she believes. So it would be really useful if there were ways, in natural language, to express such facts as that, prior to Jimmy's speaking, Lois did not have the M-belief because she did not yet have the S-belief. And it is not plausible that natural languages are so impoverished that they are unable to express such things. But the obvious way to express such a thing would be to say something like:

(B3S) Lois did not believe that the guy who can save Metropolis was at the meeting because she did not believe that Superman was at the meeting.

But Russellians hold that (B3S) expresses the same proposition as:

<sup>40</sup> That said, it is not obvious that Braun's position is really coherent. The sorts of arguments against the Naïve View that were rehearsed in Section 2, and which are the only decent arguments against it that are known to be me, depended upon the assumption that singular beliefs have explanatory work to do, and Braun now seems to be very close to denying that. But if you're prepared to accept that singular beliefs are explanatorily idle, then it is not clear to me why you need to reject the Naïve View, after all.

<sup>41</sup> Berra was a baseball player for the New York Yankees during one of their most dominant stretches and is a member of the Baseball Hall of Fame. But he is perhaps best known for such iconic remarks as "It ain't over 'til it's over", "When you get to the fork in the road, take it", and "Nobody goes there anymore, it's too crowded", which is as good an example of quantifier domain restriction as there is. The remark mentioned in the text made perfect sense in context. Berra was talking about how often Mickey Mantle and Roger Maris would hit back-to-back home runs. The first *déjà vu* was when Maris hit a home run right after Mantle had done so. The next *déjà vu* was when that happened again.



(B3C) Lois did not believe that the guy who can save Metropolis was at the meeting because she did not believe that Clark Kent was at the meeting.

And (B3C) is false, both according to Russellians and according to me. Hence, Russellians are committed to regarding (B3S) as false, in which case they must deny that (B3S) expresses that Lois lacked the M-belief because she lacked the S-belief, since that is indeed why she lacked the M-belief.

There are many other epicycles we might ride for a while. But the general strategy should now be clear. Singular beliefs, if they are to be worth having, must figure in various sorts of causal and explanatory relations. Those sorts of relations are ones in which ordinary speakers might well take an interest. That makes it a natural question how such relations might be expressed in natural language. But the semantic doctrines constitutive of Russellianism make it impossible in principle for any sentence of English ever to distinguish between Lois's having the C-belief and her having the S-belief. So Russellians are committed to there being a large expressive vacuum in English, not just in "unusually recalcitrant cases", but much more generally, since cases of the sort we have been discussing are easily constructed for any singular belief.

That is a possible view, no doubt.

## 6 Conclusion

Here, then, is the substitution argument as I have suggested we should understand it.

1. The S-belief and the C-belief are distinct, since they play different roles in Lois's psychology.
  - (a) If you don't think beliefs and the like play any role in people's psychology, then see Section 3.
2. Ordinary speakers might well take an interest in the difference between Lois's having the S-belief and her having the C-belief, so that difference should be expressible in English.
3. The only English sentences that plausibly express that Lois has the S-belief as opposed to the C-belief, or *vice versa*, are sentences like:

- (LS) Lois believes that Superman is at the school board meeting.
- (LC) Lois believes that Clark Kent is at the school board meeting.
- (a) Moreover, the fact that it is Lois's having (or not having) the S-belief that is responsible for certain of her actions, and other beliefs, ought to be expressible in English.
- (b) The only sentences that plausibly express such relations embed sentences like (LC) and (LS).
4. Since Lois can have the C-belief without having the S-belief, (LC) can be true when (LS) is false.

This argument does not depend upon any appeal to 'intuitions' that might be 'explained away' pragmatically. Facts about what ordinary speakers are inclined to say do play a role in the argument, but the objection being offered to the Russellian account of belief-ascription is *not* that it forces us to 'abandon our intuitions'. The objection is that the Russellian can offer no plausible answer to the question how the difference between the S-belief and the C-belief might be expressed in English. That ordinary speakers sometimes say things like "Lois believed that Clark was at the meeting, but not that Superman was"—namely, when they might want to distinguish between Lois's having the C-belief and her having the S-belief—is invoked in support of step (3) of the argument. That observation is not, by itself, offered in support of the claim that (LC) and (LS) have different truth-conditions. That conclusion follows only given the claim made at step (1), for which we have an independent argument.

That, then, allows us to answer Saul. The problem was supposed to be that, if we give a pragmatic account of the apparent contrast between these two sentences:

(T) Superman gets better tables than Clark Kent.

(T?) Superman gets better tables than Superman.

then we need some special reason to resist a pragmatic account of the corresponding contrast between (LS) and (LC). But there is no such correspondence. The contrast between (T) and (T?) *really is* just a matter of what people are inclined to say or, if you prefer, of 'intuition': of first-blush, naïve response. A pragmatic explanation of *those* 'intuitions' is entirely appropriate. The claim that (LC) can be true when (LS) is false, by contrast, is based upon an argument, namely, the one just

summarized. It is no response to that argument to insist that (LC) and (LS) must have the same truth-value and then to try to explain why speakers might wrongly regard them as having different truth-values. That would imply that (LS) and (LC) cannot be used to express the difference between Lois's having the S-belief and her having the C-belief, but it would leave unanswered the crucial question how that difference *can* be expressed in English.

Saul, then, is right about this much: If you think 'truth-conditional intuitions' play a significant role in the substitution argument, then you will find yourself in a bind if you are also inclined to think that the substitution argument threatens Russellian accounts of attitude ascription. That is a point well worth making. But the lesson we ought to learn, or so I am claiming, is that it is simply a mistake to think that 'intuitions' do play any significant role in the substitution argument.<sup>42</sup>

## References

- Bach, K. (2000). 'A puzzle about belief reports', in K. Jaszczolt (ed.), *The Pragmatics of Propositional Attitude Reports*. Oxford, Elsevier Science, 99–109.
- Barber, A. (2000). 'A pragmatic treatment of simple sentences', *Analysis* 60: 300–8.
- Bealer, G. (1996). 'A priori knowledge and the scope of philosophy', *Philosophical Studies* 81: 121–142.
- (1998). 'Intuition and the autonomy of philosophy', in M. R. DePaul and W. Ramsey (eds.), *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham MD, Rowman & Littlefield, 201–239.
- Braun, D. (1998). 'Understanding belief reports', *Philosophical Review* 107: 555–95.

<sup>42</sup> This material was presented in lectures at Duke University, in October 2013, and at the *Analytic Philosophy* Symposium, at the University of Texas, in December 2013. Thanks to everyone present for their questions and comments, especially Derek Ball, David Barack, David Beaver, Ray Buchanan, Josh Dever, Sinan Dogramaci, Mike Martin, Wayne Normal, Kate Ritchie, Mark Sainsbury, David Sanford, David Sosa, Walter Sinnott-Armstrong, Orlin Vakarelov, David Wong, and a few other people whose names I was not able to determine. Special thanks to Jason Stanley. But for his enthusiasm about these ideas, this paper would never have been written.

- (2001). ‘Russellianism and explanation’, *Philosophical Perspectives* 15: 253–89.
- Braun, D. and Saul, J. (2002). ‘Simple sentences, substitutions, and mistaken evaluations’, *Philosophical Studies* 111: 1–41.
- Camp, E. (2006). ‘Metaphor and that certain “je ne sais quoi”’, *Philosophical Studies* 129: 1–25.
- Cappelen, H. (2012). *Philosophy Without Intuitions*. Oxford, Oxford University Press.
- Cappelen, H. and Dever, J. (2014). *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford, Oxford University Press.
- Carey, S. (2009). *The Origin of Concepts*. New York, Oxford University Press.
- Fodor, J. (1994). *The Elm and the Expert*. Cambridge MA, MIT Press.
- Forbes, G. (1997). ‘How much substitutivity?’, *Analysis* 57: 109–13.
- (1999). ‘Enlightened semantics for simple sentences’, *Analysis* 59: 86–91.
- Frege, G. (1984a). *Collected Papers on Mathematics, Logic, and Philosophy*, McGuinness, 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.
- Gettier, E. (1963). ‘Is justified true belief knowledge?’, *Analysis* 23: 121–3.
- Goldman, A. (2012). ‘Theory of mind’, in E. Margolis, *et al.* (eds.), *The Oxford Handbook of Philosophy of Cognitive Science*. Oxford, Oxford University Press, 402–24.
- Gopnik, A. (2009). *The Philosophical Baby: What Children’s Minds Tell Us About Truth, Love, and the Meaning of Life*. New York, Picador.

- Grice, H. P. (1989). 'Logic and conversation', in *Studies in the Ways of Words*. Cambridge MA, Harvard University Press, 22–40.
- Heck, R. G. (1995). 'The sense of communication', *Mind* 104: 79–106.
- (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.
- (2012). 'Solving Frege's puzzle', *Journal of Philosophy* 109: 132–74.
- Ludlow, P. (2011). *The Philosophy of Generative Linguistics*. Oxford, Oxford University Press.
- Moore, J. (1999). 'Saving substitutivity in simple sentences', *Analysis* 59: 91–105.
- Nagel, J. (2012). 'Intuitions and experiments: A defense of the case method in epistemology', *Philosophy and Phenomenological Research* 85: 495–527.
- Partee, B. (1988). 'Semantic facts and psychological facts', *Mind and Language* 3: 43–52.
- Perry, J. (1993). 'The problem of the essential indexical', in *The Problem of the Essential Indexical, and Other Essays*. New York, Oxford University Press, 33–49.
- Pitt, D. (2001). 'Alter egos and their names', *Journal of Philosophy* 98: 531–52.
- Predelli, S. (1999). 'Saul, Salmon, and Superman', *Analysis* 59: 113–6.
- (2004). 'Superheroes and their names', *American Philosophical Quarterly* 41: 107–23.
- Recanati, F. (1993). *Direct Reference: From Language to Thought*. Oxford, Basil Blackwell.
- (2013). *Mental Files*. Oxford, Oxford University Press.

- Salmon, N. (1986). *Frege's Puzzle*. Cambridge MA, MIT Press.
- Saul, J. (1997). 'Substitution and simple sentences', *Analysis* 57: 102–8.
- (1998). 'The pragmatics of attitude ascription', *Philosophical Studies* 92: 363–89.
- (2007). *Simple Sentences, Substitution, and Intuitions*. Oxford, Oxford University Press.
- Schneider, S. (2005). 'Direct reference, psychological explanation, and Frege cases', *Mind and Language* 20: 423–47.
- (2011). *The Language of Thought: A New Philosophical Direction*. Cambridge MA, MIT Press.
- Searle, J. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, Cambridge University Press.
- Segal, G. (1997). 'Content and computation: Chasing the arrows', *Mind and Language* 12: 490–501.
- (2000). 'Frege's puzzle as some problems in science', *Rivista di Linguistica* 8: 375–388.
- Soames, S. (1987). 'Substitutivity', in J. J. Thomson (ed.), *On Being and Saying: Essays for Richard Cartwright*. Cambridge, MIT Press, 99–132.
- (1995). 'Beyond singular propositions?', *Canadian Journal of Philosophy* 25: 515–49.
- Stanley, J. and Szabó, Z. G. (2000). 'On quantifier domain restriction', *Mind and Language* 15: 219–261.
- Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge MA, MIT Press.
- Williamson, T. (2008). *The Philosophy of Philosophy*. Oxford, Oxford University Press.