Three Problems for Richard's Theory of Belief Ascription*

Theodore Sider *Canadian Journal of Philosophy* 25 (1995): 487–514)

Some contemporary Russellians, defenders of the view that the semantic content of a proper name, demonstrative or indexical is simply its referent, are prepared to accept that view's most infamous apparent consequence: that coreferential names, demonstratives, indexicals, etc. are intersubstitutable *salva veritate*, even in intentional contexts. Nathan Salmon and Scott Soames argue that our recalcitrant intuitions with respect to the famous apparent counterexamples are not semantic intuitions, but rather pragmatic intuitions. Strictly and literally speaking, Lois Lane believes, and even *knows* that Clark Kent is identical to Superman, since she believes and knows that Superman is identical to Superman. Salmon and Soames attempt to soften our reaction to this shocker by allowing that it is typically *misleading* to utter the sentence 'Lois Lane knows that Clark Kent is identical to Superman', since it pragmatically implicates, without semantically entailing, that Lois Lane would accept the sentence 'Clark Kent is identical to Superman'. Our compulsive tendency to claim that 'Lois Lane knows that Clark Kent is Superman' is *false*, rather than merely misleading, is due to a confusion between semantics and pragmatics, between truth conditions and conditions of appropriateness of utterance.[1]

It is probably fair to say that the common reaction to this move in defense of Russellianism is negative. Mark Richard says the following:[2]

> …other than using bribery, threats, hypnosis, or the like, there is simply nothing you can do to get most people to say that Jones believes that Tully was an orator, once they know that Jones sincerely denies 'Tully was an orator', understands it, and acts on his denial in ways appropriate thereto. In particular, pointing out that Jones can express something he believes with 'Cicero was an orator' seems simply irrelevant to most people… The Russellian is correct when he says that our intuitions about truth conditions are not wholly reliable. But they are certainly not to be *ignored*.

It is indeed hard to accept that Lois Lane believes that Clark Kent is Superman; it would be nicer to be able to accept what is attractive about Russellianism

*I would like to thank Mark Aronszajn, David Cowles, Ed Gettier, two anonymous referees, and especially David Braun and Mark Richard for their help with this paper.

[1] See Salmon (1986); Soames (1987, 218–20) and Grice (1975) on pragmatic implicature. Kaplan (1989) and Kripke (1972) are classic works associated with Russellianism.

[2] Richard (1990, 125).

without paying the high price that Salmon and Soames are willing to pay. In his recent book, *Propositional Attitudes*, Richard offers a theory that promises to do just this.[3]

I believe that Richard's view incorporates many features of an ideal theory of belief ascription. But I also believe it faces difficulties. I will present three objections to Richard's theory. The first two objections function as a unit: it is possible to modify Richard's view to circumvent each of the first two objections taken individually, but combining the two modifications leads to trouble. The final objection is a "logical" objection that stands on its own, and raises issues of general interest in the debate between Russellianism and its rivals. I will first present Richard's theory, and then turn to the objections.

1. Richard's Theory

The core idea behind Richard's theory is an attractive one. When I use a belief ascription sentence such as "Lois Lane believes that Clark Kent can fly", according to Richard I am making a claim about what sentences Lois "accepts", or has in her "representational system", and in particular claiming that she accepts some sentence that is *correctly* represented by "Clark Kent can Fly".[4] Lois *does* accept "Superman can fly", but if this sentence cannot be correctly represented by "Clark Kent can fly", then the belief ascription sentence is false. What is novel about Richard's theory is that what counts as a correct representation of a sentence may vary from one context of utterance to another. Thus, the truth value of utterances of "Lois believes that Clark Kent can fly" may vary from one context of utterance to another, without any variation in the facts about Lois; this would happen if the contexts differed over what "Clark Kent can fly" may correctly represent. Analogously, the truth values of utterances of "this table is flat" may vary from one context to another without the table itself altering, because the standards governing what it is to count as "flat" may vary between those contexts.

For a more detailed presentation it will be necessary to introduce the entities that serve as Richard's propositions: "Russellian annotated matrices" (RAMs).

[3]See chapter 3.

[4]The terminology here should not be taken to imply that the sentences accepted must be natural language sentences, or that the notion of acceptance here is exactly the everyday notion of acceptance; these are technical notions that Richard develops in his own way. See Richard (1990, chapter 3), especially pp. 181–90.

We can represent an English sentence like 'Twain is happy' with the following ordered pair:

$\langle$'is happy', 'Twain'$\rangle$

The Russellian associates the proposition that Twain is happy with the pair:

$\langle$**happiness**, Twain$\rangle$

A RAM, on the other hand, combines features of sentences and Russellian propositions. First, we pair phrases like 'is happy' and 'Twain' with their Russellian referents to get *annotations*. The following are annotations:

$\langle$'is happy', **happiness**$\rangle$

$\langle$'Twain', Twain$\rangle$

Finally, we pair annotations together to get RAMs. Thus, the RAM determined by 'Twain is happy' is:

$\langle\langle$'is happy', **happiness**$\rangle$, $\langle$'Twain', Twain$\rangle\rangle$

Note that this RAM is distinct from the RAM determined by 'Clemens is happy':

$\langle\langle$'is happy', **happiness**$\rangle$, $\langle$'Clemens', Twain$\rangle\rangle$

A person, S, accepts some sentences. These sentences determine a set of RAMs called S's "representational system" (RS).[5] Think of S's "RS" as being determined by the sentences that S has "written in her head". Likewise, in the belief ascription sentence $\ulcorner S$ believes that $\phi\urcorner$, the component sentence ϕ determines a RAM. Richard wants to say that the belief ascription sentence, as uttered in a certain context, c, is true just in case the RAM determined by ϕ is an appropriate-in-context-c representation of some RAM in S's RS.

More carefully[6], a *correlation function* is defined as a function that maps annotations to annotations and preserves reference. So a correlation could map

[5] In this paper I will pretend that the "linguistic parts" of RAMs in one's RS are pieces of natural language. This ignores some complications present in Richard's final view—see Richard (1990, 181–90).

[6] See Richard (1990, 136 ff.).

$\langle$'Londres', London$\rangle$ to $\langle$'London', London$\rangle$, but not to $\langle$'Paris', Paris$\rangle$. Say that one RAM p *represents* another RAM q *under* correlation f ("Rep(p,q,f)") iff every annotation in p is in the range of f, and q results from p by replacing every annotation in p by its image under f. (I will sometimes talk loosely of one *sentence* representing another.) A *restriction* is a triple $\langle x,a,S\rangle$, where x is a person, a is an annotation, and S is a set of annotations; restriction $\langle x,a,S\rangle$ is *relevant* to person u iff $u=x$. Intuitively, a restriction $\langle x,a,S\rangle$ says that when talking about x's RS, we must use annotation a only to represent annotations in S. Where s is a set of restrictions, f a correlation, and u a person, we say "Obey(s,f,u)" iff for every restriction $\langle u,a,S\rangle \in s, f(a)\in S$.

According to Richard, every possible context of utterance c provides a set $r(c)$ of restrictions; these restrictions are typically the result of the shared intentions of those involved in the conversation. A belief sentence is true, as uttered in a context, iff the RAM determined by the subordinate clause represents some RAM in the believer's RS under some correlation that obeys the restrictions in the context relevant to the believer:

> $\ulcorner t$ believes that $\phi\urcorner$ is true in c iff $\exists f \exists q \exists x[x$ is the referent of $t \wedge q \in$ x's RS $\wedge$ Obey$(r(c),f,x) \wedge$ Rep(the RAM determined by $\phi,q,f)]$

This is best illustrated by means of an example. Suppose two people are having a conversation about their friend Hank. They know that he thinks that Twain is a novelist, but they wonder whether he thinks that Clemens is a novelist. Maybe they see him looking at a magazine article that mentions only 'Clemens'. In this context, the conversationalists are focusing on Hank's usage of 'Twain' and 'Clemens'. When they use the term 'Twain' in discussing Hank's beliefs, they intend this to represent Hank's usage of this very term—'Clemens' won't do. Similarly, they want to use 'Clemens' only to represent Clemens. They *know* that Hank accepts the sentence 'Twain is a novelist'; their only question is whether he accepts 'Clemens is a novelist' as well—thus, the fact that Hank accepts the first sentence shouldn't be sufficient for the truth of 'Hank thinks that Clemens is a novelist'.

This is brought about formally by the following restriction being operative in the context:

> $\langle$Hank, $\langle$'Clemens', Twain$\rangle$, $\{\langle$'Clemens', Twain$\rangle\}\rangle$

We may abbreviate this as follows:

Hank: 'Clemens' $\rightarrow$ {'Clemens'}

Intuitively, what this says is that in talking about Hank's beliefs, the word 'Clemens' may be used only to represent occurrences of 'Clemens' in Hank's RS; in particular, 'Clemens' should not be used to represent 'Twain'. For if Hank does not accept the sentence 'Clemens is a novelist', then

(o) Hank believes that Clemens is a novelist

should be false in this context, despite the fact that the RAM determined by 'Twain is a novelist' is in Hank's RS. But we can imagine other contexts of utterance in which (o) ought to be true, without altering Hank in any way. Our two conversationalists might discuss Hank's beliefs about Clemens without caring how Hank would express those beliefs. Perhaps they shift their attention from the magazine Hank is reading, and instead count how many American novelists Hank thinks there are. They say "well, Hank knows about Faulkner, Hemingway, and Clemens—that makes three." In such a context, (o) ought to be true, despite the fact that Hank doesn't accept 'Clemens is a novelist'. Richard accounts for this by saying that in the new context, since the conversationalists are no longer focusing on what particular name Hank associates with Clemens, there are no restrictions on what we may use 'Clemens' to represent; any name for Clemens will do.

Richard has a straightforward solution to Frege's puzzle of Hesperus and Phosphorus. Any rational person, S, who has heard of Hesperus and Phosphorus knows that Hesperus is Hesperus and Phosphorus is Phosphorus, and so S accepts sentences of the form $\ulcorner H$ is $H \urcorner$ and $\ulcorner P$ is $P \urcorner$, where H and P are names that refer to Venus. But from this it does not follow that S believes that Hesperus is Phosphorus. For i) S may not accept the sentence $\ulcorner H$ is $P \urcorner$, and if not, and if ii) the current context of utterance determined by my writing this paragraph and your reading it creates the restrictions:

S: 'Hesperus' $\rightarrow$ {'H'}

S: 'Phosphorus' $\rightarrow$ {'P'}

then the sentence 'S believes that Hesperus is Phosphorus' comes out false (as uttered in the current context). Richard's theory also handles other traditional belief puzzles quite nicely. The idea that 'believes' is indexical seems to explain the fact that people have wildly varying intuitions about the truth of belief ascription sentences in certain cases, depending on the way the cases are presented. This is especially evident in Kripke's example of puzzling Pierre, in which we are drawn to call the same sentence both true and false in different contexts.[7]

There are two principle virtues of Richard's theory. The first is that it has powerful resources for accounting for the strong intuitions of speakers in the various famous puzzle cases in the philosophy of language.[8] Unlike Salmon and Soames, Richard can take our intuitions that coreferential proper names cannot be substituted *salva veritate* at face value, for he can say that those intuitions apply to contexts in which there are appropriate restrictions applying to the names in question. Secondly, Richard's theory is consistent with the rejection of a descriptivist theory of names. Failure of substitutivity of coreferential names, as we have seen, does not derive from differences in semantic content between such names, but rather from differing contextual restrictions on representation. Richard can thereby join Salmon and Soames in embracing the direct reference theory of names, demonstratives, and indexicals. There are theories that have the first virtue, and theories that have the second, but few claim to combine both. Thus, Richard's view takes on considerable interest, and is worthy of serious scrutiny.

2. The First Objection

In the present section I will present a fairly intricate example, which consists of two puzzles of the familiar mistaken identity type, one inside the other. It is a variant of one of Richard's examples.[9] Suppose Charlie and I discuss the beliefs of various of my students, who are taking an exam in an adjacent room. Our conversation concerns in part Odile, and her answers to test questions. We wonder whether the following is true:

(1) Odile believes that Twain is dead

The test concerns various novelists; one of the questions is 'Is Twain dead?'. In wondering about (1), we are focusing on whether Odile will answer this question correctly. In fact Odile answers *yes* to this question, so (1) seems true.

[7] See Kripke (1979) and Richard's discussion in Richard (1990, 179–80).

[8] See Richard (1990, chapter 3). Another "contextualist" theory which shares many attractive features with Richard's is that defended in Crimmins and Perry (1989).

[9] See Richard (1990, chapter 3).

Another of the questions on the exam is: 'Is Twain famous?'. Unfortunately, Odile has not studied well enough; she thinks that 'Twain' refers to an obscure dead author. So she answers *no*. So if in this context we were to utter:

(2) Odile doesn't believe that Twain is famous

it would seem to be true. However, we may also stipulate that Odile *would* assent to '*Clemens* is famous'—the RAM determined by this sentence is in her RS. This means that 'Twain' in (2) had better not represent Odile's use of 'Clemens'; otherwise (2) would turn out false. So it seems that the restrictions in the context don't allow us to represent occurrences of 'Clemens' in Odile's RS with 'Twain'.

We can extend the example further so that the context forbids our using 'Twain' to represent other terms Odile uses to refer to Twain. Suppose the room contains a large picture of Twain to which Odile sometimes gestures. She knows that the man in the picture is a famous author, but she does not know that he is named 'Twain'. Odile would accept 'he is famous' while pointing at the picture. If we allowed 'Twain' in (2) to represent Odile's uses of 'he', again (2) would turn out false. Putting all this together, it seems that the following restriction is in effect:

R_1: Odile: 'Twain' $\rightarrow$ {'Twain'}

R_1 requires that we use 'Twain' only to represent occurrences of 'Twain' in Odile's RS.

In addition to discussing Odile, Charlie and I also discuss the beliefs of Amanda. We believe Amanda to be an excellent student—probably, we say, she has finished her exam already, getting all the answers correct, and has now become engrossed with the picture of Twain on the wall. The picture is her favorite; she has spent many hours admiring it. She has read many times the long caption describing Twain's exploits (though the caption never mentions 'Twain'). The caption fails, however, to mention Twain's death date, so Amanda believes that the man in the picture is still alive. Amanda refers to the man in the picture using phrases like 'the man in the picture', 'he', etc. Call these Amanda's "perceptual Twain-terms". In our conversation Charlie and I have been focusing on Amanda's perceptual Twain-terms. I have been asking Charlie about what Amanda thinks about Twain, the man in the picture, and Charlie's answer is "she thinks Twain is a famous author, but she doesn't think he's dead". Charlie takes the following to be true:[10]

(3) Amanda doesn't believe that Twain is dead.

For when talking about Amanda, we aren't focusing on the fact that we think she answers *yes* to the test question 'Is Twain dead?' (we think she got all the answers right, recall); rather, we are focusing on the fact that she would reject the sentence "he is dead" while pointing at the picture of Twain. So an additional restriction seems to be present:

R_2: Amanda: 'Twain' $\rightarrow S$

where S is the set of Amanda's perceptual Twain-terms (e.g. 'he', 'that guy in the picture'). S must *not* contain 'Twain', for Charlie believes (3) to be true even though he thinks that Amanda assented to 'Twain is dead' when writing her exam.

There is one final twist to the story. Unbeknownst to Charlie and me, Odile *is* Amanda. We think that there are two people in the other room, one taking a test and the other gazing at the picture. But there is only one. Odile—Amanda—took the test, and got most of the answers right (except for the one about Twain's being famous). And the picture on the wall is indeed her favorite picture; she thinks it is the picture of a famous living author.

But this means trouble. Since Amanda and Odile are one and the same person, we may rewrite R_1 as follows:

R_3: Amanda: 'Twain' $\rightarrow$ {'Twain'}

After all, a restriction is defined by Richard to be an ordered triple $\langle x, a, S\rangle$, x a person, a an annotation, and S a set of annotations. So if we have two restrictions $\langle x, a, S\rangle$ and $\langle y, a, S\rangle$ where x is y, then the "two" restrictions are one. Thus, restriction R_1 is identical to restriction R_3.

R_2 and R_3 together require any correlation relevant to Amanda both to map 'Twain' to 'Twain' and to map 'Twain' to some member of S. Since 'Twain' is

[10] I'm not claiming that (3) *is* true; as will be seen below, since Odile = Amanda and (1) seems true, it's plausible that (3) is false. My claim is simply that Charlie and my intentions are sufficient for restriction R_2's being operative; after all, we think we're talking about two different people, and if we were talking about two different people, (3) would be true and R_2 would be in place.

not a member of S, no correlation satisfies both constraints. But this has drastic results. On Richard's view, (1) is true in the context if and only if there is some correlation function that obeys all of the restrictions in the context relevant to Amanda under which the RAM determined by 'Twain is dead' represents one of the RAMs in Amanda's RS. But *no* correlation function relevant to Amanda obeys all of the restrictions in the context. So (1) isn't true after all. Moreover, *no* sentence attributing a belief to Amanda (= Odile) will be true in that context, for the same reason: no correlation function obeys both R_2 and R_3. 'Amanda believes that Twain is Twain' turns out false. 'Amanda believes that the sky is blue' turns out false. These results are clearly unacceptable.

One might reply that we have here *two* contexts, rather than one; one is governed by R_2, the other by R_3. For to generate the problem I need a single context in which *both* R_2 and R_3 are operative. But this reply is unsuccessful. We may stipulate that Charlie and I have a single unified conversation about Odile and Amanda. The story could be fleshed out in such a way that we intermingle talk of the "two" students, utter conditionals with (1) as antecedent and (3) as consequent, etc. And we never intend to change the way we represent anyone's beliefs. Provided there are at least sometimes contexts in which the beliefs of two different people are discussed, it seems hard to exclude this case as such a context, for the only facts that could disqualify it are "inaccessible" to Charlie and me.

It might be replied that the context containing these conflicting restrictions is "defective" in some way.[11] Such contexts might be compared with contexts in which a contextually determined parameter is inadequately determined. Suppose, for example, that I point and say "That apple is red", but fail to demonstrate anything with 'that' because (unbeknownst to me) there are two apples in the direction I'm pointing. Or suppose that I say "The table is flat", but no determinate standards of flatness have been set up in the context, perhaps because previously in the conversation my audience and I have been shifting between different standards of precision, with no apparent pattern to the shifting.[12] In such defective contexts, it is natural to expect the utterances to lack truth value, for there seems to be no one proposition expressed. The moral seems to be that if context determines a parameter which in turn determines the proposition expressed, and no one parameter is selected, then we should expect there to be no one proposition expressed, and thus expect indeterminacy.

[11] I thank Mark Richard and an anonymous referee for helpful discussions on this point.

[12] See Lewis (1979, 245–6) on contextual determination of standards of precision.

The attempt to assimilate my problem case to uncontroversial cases of indeterminacy might proceed as follows. First, we restrict Richard's theory as originally stated to non-defective contexts—contexts without incompatible restrictions. As for the defective contexts, we may notice that any such context may be "turned into" a non-defective context simply by removing a restriction. In my example above, one may simply remove either restriction R_2 or restriction R_3. The propositions expressed by belief ascription sentences, relative to either of these non-defective "resolutions" of the original defective context, may be likened to the various propositions that would be expressed by 'The table is flat' in the example of the preceding paragraph, when we arbitrarily select one of the standards of precision between which my audience and I were shifting in the original, defective context. We may say that a sentence, S, in a defective context, c, is indeterminate; specifically, it is indeterminate over the set of propositions expressed by S in the various non-defective resolutions of c. If S comes out true in all of c's resolutions, then we may say that it is true in c; if S comes out false in all such resolutions then S is false in c; but if S comes out true in some resolutions, and false in others, then it is neither true nor false in c.

I find this response implausible, for there are crucial differences between my example and the cases of uncontroversial indeterminacy. In the uncontroversial cases, speakers' intentions and behavior are too *unspecific* to resolve the relevant parameter, whereas this is not true in my case. In the standards of precision case, for example, the participants in the conversation simply have not decided on what standards of precision are relevant, not even implicitly. If this lack of specificity were pointed out to them, they could resolve it. So it seems natural to accept indeterminacy in these cases: the intentions of the participants in the conversation are not determinate enough to single out a unique proposition expressed. But in my example, the speakers are perfectly specific in what they intend and do. Thus, it seems implausible to claim indeterminacy in this case; at the very least, the indeterminacy has no precedent in the cases of pointing and standards of precision.

Apparently, *any* plausible semantic theory will attribute indeterminacy to the cases of pointing and standards of precision.[13] And it is quite natural, pre-theoretically, to view these as cases of indeterminacy. Not so for the case of Odile; to the unprejudiced mind, surely the speakers in that case determinately

[13] Perhaps a description theory of demonstratives would avoid this consequence in the case of pointing.

assert things about Odile's beliefs, just as in more traditional cases of mistaken identity, people who are confused about identity statements may still succeed in making determinate assertions about the relevant objects, have determinate beliefs about them, etc. And it is not the case that all theories of belief ascription must appeal to indeterminacy here; even a contextualist theory of belief ascription need not make such an appeal, as I show in the next section. Thus, the appeal to indeterminacy is unmotivated and implausible.

3. A Solution: Permissions

One diagnosis of the problem is that Richard incorporates contextual factors as *restrictions*. When restrictions conflict, the number of allowable correlations decreases, sometimes down to zero. Perhaps the number of acceptable correlations should increase in such situations. This result may be achieved by using *permissions* instead of restrictions. On Richard's original view, a correlation is "innocent until proven guilty" (by violating a restriction). Restrictions rule out correlations; as the restrictions grow, the correlations dwindle. We might instead take correlations to be "guilty until proven innocent" (by being permitted by some permission). As the permissions grow, the allowed correlations will grow as well.

Call a "permission" any triple $\langle x, a, S\rangle$, x a person, a an annotation, and S a set of annotations. Where s is a set of permissions, f a correlation, and u a person, we write "Permit(s, f, u)" just when, for every annotation a in f's range, there is some permission $\langle u, a, S\rangle$ in s such that $f(a) \in S$. The other definitions stay the same.

Think of each context c as providing a set $p(c)$ of permissions for that context. Permissions will be the mirror images of restrictions. In a context where Richard had no restrictions, there will need to be many permissions, one for every annotation that could possibly be used to represent an annotation in someone's RS. In a context with many restrictions, there will need to be few permissions.

On this new theory, the account of belief attribution is as follows:

> $\ulcorner t$ believes that $\phi\urcorner$ is true in c iff $\exists f \exists q \exists x[x$ is the referent of $t \wedge q \in$ x's RS $\wedge$ Permit($p(c), f, x$) $\wedge$ Rep(the RAM determined by ϕ, q, f)]

As an illustration of this view, consider its application to the last story of the previous section. The crucial sentences were:

(1) Odile believes that Twain is dead

(3) Amanda doesn't believe that Twain is dead

(1) seems true. So one permission operative in the context seems to be:

P_1: Odile: 'Twain' → {'Twain'}

But we need another permission. On the new view, correlations are guilty until proven innocent. If P_1 were the only permission, then (1) would be false, since no correlation that mapped 'is dead' to anything would be permitted. So the following permission must be present as well:

P_2: Odile: 'is dead' → {'is dead'}

There will need to be many other permissions as well. Surely 'is dead' could represent other synonymous phrases. More importantly, permissions for almost all words must be present. 'is famous' ought to be able to represent 'is famous', 'is happy' should be able to represent 'is happy', 'Charlie' should be able to represent 'Charlie', etc. Moreover, since Richard allows no restrictions on variables, given any assignment to the variables, every context will need to contain permissions allowing variables to represent any coreferential terms whatsoever.[14]

There will need to be still other permissions in the context, due to our conversation that involves (3). Recall that we think that Amanda is looking at the picture of Twain and thinking to herself "he is a famous living author", and we attribute beliefs to her about Twain, using 'Twain' to represent her uses of 'he', 'that guy', etc. It seems that P_3 is an operative permission as well as P_1 and P_2:

P_3: Amanda: 'Twain' → S

where *S* is, as before, the set of Amanda's perceptual Twain-terms.

Now, since Odile *is* Amanda, just as R_1 was identical to R_3 above, P_1 is identical to P_4:

P_4: Amanda: 'Twain' → 'Twain'

[14]Richard (1990, 151–3).

When we had restrictions instead of permissions, no correlations were allowed since no correlation obeyed all the restrictions in the context. But there is no corresponding problem here for the new theory. Since both P_3 and P_4 are present, correlations that map 'Twain' to 'Twain' are permitted, *as well as* correlations that map 'Twain' to members of S. So we do not have the damaging result that no correlations are allowed in the context.

In cases like the one we have been discussing, the new theory frees up unexpected correlations. For example, the new theory has the result that (3) is false in the context, since we are permitted by P_1 to represent occurrences of 'Twain' in Odile's RS by 'Twain'. But this might be thought to be no defect. That same permission has the result that (1) is true in the context. But the sentence

(4) Amanda believes that Twain is dead

results from (1) by substituting 'Amanda' for 'Odile'. Since these two terms are coreferential, and since the substitution takes place on the *left-hand* side of 'believes', we might expect (4) to be true in the context, and hence expect (3) to be false. (More on this in the next section.)

One might worry that this new theory gets us out of the frying pan only to land us in the fire. Notice that the sentence:

(4′) Amanda believes that Twain is not dead

also turns out true. For P_3 allows us to use 'Twain' to represent occurrences of members of S (such as 'he' when accompanied by her pointing to the picture of Twain) in Amanda's RS. Since she would accept 'he is not dead' when pointing at the picture, (4') is true. Doesn't the fact that (4') and (4) are both true mean that Amanda is in some sense irrational? Following Kripke, we can stipulate that Amanda is a leading logician; she would *never* "let contradictory beliefs pass".[15]

We could reply as follows. The natural way to interpret a dialogue containing (4) and (4′) is to evaluate those two sentences according to a single correlation function. But there is no one correlation that makes both true. A correlation is a function, and so must map 'Twain' to some term T. But being a wise logician, for no term T does Amanda accept two sentences of the form:

[15] The phrase is from Kripke (1979, 122).

T is dead

T is not dead

Since there is no correlation that makes both (4) and (4′) true, the natural way of understanding a dialogue containing both sentences makes the combination of those sentences false. And since the natural way of interpreting the combination of (4) and (4′) makes them false, we need not conclude that Amanda is irrational.

Richard takes this approach to Kripke's Pierre puzzle. Indeed, he argues that the most natural way to handle multiple ascriptions is often to evaluate all of the belief ascriptions according to a single correlation function. This is somewhat like treating the multiple ascriptions as ascriptions of belief in a long conjunction. The sentences ascribing beliefs in the conjuncts do not have any meaningful truth value assignment in isolation. If he is right about this, then we must evaluate (4) and (4′) jointly, and since there is no single correlation in virtue of which (4) and (4′) turn out true, they jointly turn out false.[16] For the sake of argument, let us grant Richard this sort of reply.

4. An Alternate Solution: Keying Restrictions to Words

As an aside, I want to mention an entirely different solution to the problem of section 2. We could key restrictions to words and not to people. In the example of story 4, the problem with Richard's original theory was that no correlation could obey both R_1 and R_2:

R_1: Odile: 'Twain' → {'Twain'}

R_2: Amanda: 'Twain' → S

since Odile = Amanda. But in their places we could put

R'_1: 'Odile': 'Twain' → {'Twain'}.

R'_2: 'Amanda': 'Twain' → S

Since 'Odile' ≠ 'Amanda', there is no problem with correlations obeying both restrictions.

On the new theory, a restriction would be a triple $\langle \lambda, a, S\rangle$, where λ is a pair $\langle t, x\rangle$ (t being a term that refers to person x), a is an annotation, and S is a set of annotations. We would say "Obey(s, f, λ)" iff for every restriction $\langle \lambda, a, S\rangle \in s, f(a) \in S$. The theory of belief ascription would be:

[16] See Richard (1990, 173–80).

$\ulcorner t$ believes that $\phi\urcorner$ is true in c iff $\exists f \exists q \exists x[x$ is the referent of $t \wedge q \in$ x's RS $\wedge \text{Obey}(r(c), f, \langle t, x\rangle) \wedge \text{Rep}($the RAM determined by $\phi, q, f)]$

Recall that the permissions theory of section 3 freed up unexpected correlations in our example. In particular, we noted that:

(4) Amanda believes that Twain is dead

turned out true, even though we intend to be focusing on Amanda's attitudes towards the sentence 'He is dead' (which she rejects). Moreover, since (4′) turned out true as well, that theory seemed to imply that Amanda was somehow irrational. If you didn't like the response to this problem I considered at the end of that section, you may prefer the theory of the current section, for it does not imply that (4) is true.

But this advantage comes with a certain price: truth is sometimes not preserved through substitution of coreferential terms *outside* of belief operators. On the present theory, (4) is false, even though

(1) Odile believes that Twain is dead

is true. But Odile *is* Amanda. So the question of the proper way to fix Richard's theory reduces to the question: can we substitute coreferential terms in the subject position of belief sentences? If we can, then the permissions theory seems to be vindicated. If not, then we might want to consider the theory of the present section. I think it is clear that the permissions theory is superior. On the theory of the present section, the following sentence turns out true in the context of Charlie and my discussion of Odile and Amanda:[17]

Odile = Amanda, and Odile believes that Twain is dead, and Amanda does not believe that Twain is dead.

But one presumably ought to be able to infer from this that:

$\exists x \exists y(x = y \wedge x$ believes that Twain is dead $\wedge\, y$ does not believe that Twain is dead)

[17] One might object here along the lines of the objection at the end of section three: there is no *single* correlation function that makes the latter two conjuncts true, but the natural way to interpret this sentence is indeed in terms of a single correlation function. But I disagree that the natural way to interpret this sentence is according to a single correlation function, since the speakers in the dialogue don't know that Odile is identical to my student.

but from this it seems to follow that:

> $\exists x(x$ believes that Twain is dead $\wedge\, x$ does not believe that Twain is dead)

which seems necessarily false. I will therefore assume that the permissions solution is preferable to that of keying restrictions to words rather than persons.[18]

5. The Second Objection

Let us return to Richard's original (restriction-based) theory for the second objection (we will return to the permissions theory in section 7 below). An advantage of Richard's view is its flexibility. Richard accommodates our intuitions that sometimes substitution of co-referential terms on the *right-hand* side of belief sentences is invalid—in such examples (like the example of Lois Lane at the beginning of the paper) our intuitions conflict with the theories of Salmon and Soames. But in other cases, our talk about beliefs seems to match Salmon and Soames's predictions. In these cases, we do not care how the believer would express his or her belief—we care only about the Russellian proposition believed. An attractive feature of Richard's theory is that it can accommodate these examples as well: they involve contexts with no restrictions. Suppose Odile and I are discussing our mutual friend Mark. Out of the blue, we wonder "Does Mark think that Laurie is famous?". Laurie is another mutual friend of ours, but we have no idea whether Mark has ever met Laurie, or in what circumstances he might have met her. We are prepared to count 'Mark believes Laurie is famous' true so long as Mark accepts some sentence of the form $\ulcorner T$ is famous$\urcorner$ where T is some name or demonstrative that refers to Laurie. This would be an example of a context without restrictions on Mark's terms that refer to Laurie.

But there is a certain kind of flexibility in this area that Richard's theory lacks, which is due to the fact that Richard requires correlations to be *functions*.[19] Let terms 'a' and 'b' be coreferential. On Richard's view, the RAM determined

[18] This argument will be blocked at some point by the theory of the present section, depending on how quantification into the left-hand position of belief sentences is handled. The objection is just that the argument *shouldn't* be blocked—it seems valid.

[19] This point has also been made (independently) by Crimmins (1992, 192). Richard replies in Richard (1993, 127–9). Whatever the merits of Richard's reply, it does not apply to my examples in this section, not directly anyway.

by 'Raa' cannot represent the RAM determined by 'Rab' under any correlation, since a correlation, being a function, cannot map 'a' to both 'a' and 'b'. Similarly, the RAM determined by '$Fa \wedge Ga$' cannot represent the RAM determined by '$Fa \wedge Gb$', nor can the RAM determined by '$Fa \wedge Fb$' represent the RAM determined by '$Fa \wedge Gb$'.

I think this inflexibility is unwelcome. Suppose Jane is going blind, and I am a doctor testing her sight. I first show her a ball I call "*A*"; she is to tell me its size. She says 'large'. Later, I show her a ball that I call "*B*". She must tell me its color. 'Red', she says. Though she does not know it, I showed her the same large red ball each time. Ball *A* and ball *B* are identical.

Later, I discuss her performance with my fellow doctors. She did well, I say. She knew that ball *A* was large and red. Our conversation centers on whether Jane can recognize large red things when she sees them; we don't care what name Jane associates with the ball, nor whether she associates the same name with the red ball and the large ball. In this context, my utterance of

(5) Jane believes that *A* is large and *A* is red

should turn out true. But, since Jane doesn't know that ball $A =$ ball B, she doesn't accept '*A* is large and *A* is red'. What she accepts is '*A* is large and *B* is red'. But the first sentence cannot represent the second under any correlation, since such a correlation would need to map '*A*' to '*A*' and also to '*B*'. Hence (5) isn't true, on Richard's view.[20]

In an appropriate context, the sentence '*A* is large and *B* is red' could represent '*A* is large and *A* is red', since a correlation function could map both '*A*' and '*B*' to '*A*'. It is only the reverse representation that is not allowed. So, if Jane had uttered '*A* is large and *A* is red', given an appropriate context we could report her belief using the sentence

(6) Jane believes that *A* is large and *B* is red

But if this is so, then surely (5) should be capable of expressing a truth in an appropriate context. It is this asymmetry between (5) and (6) to which I object.

[20]Incidentally, Richard's theory also has the result that 'Jane believes that *A* is large and red' turns out false in every context, whereas it would be natural to count it true in the context I discuss in the text. Moreover, it is hard to see how Richard's theory could be revised to respect this intuition.

The situation can arise with predicates as well as singular terms. Suppose that John is unaware that 'groundhog' and 'woodchuck' are coreferential. Despite this ignorance, John is a professional animal sorter, who travels from place to place sorting animals. In one place his instructions are to sort the groundhogs from the raccoons. In this he succeeds. In another place he must sort the woodchucks from the squirrels. His performance again is stellar. Suppose you and I discuss John's doings. We focus on his ability in each case to tell the groundhogs from the other animals—in the first case from raccoons, and in the second place from squirrels. I remember in particular two groundhogs; one from each place. The first is named Jerry, and the second is named Jerome. I remark that John knew that Jerry and Jerome were groundhogs without a moment's hesitation. The following sentence ought to be true in our context:

(7) John believes that Jerry is a groundhog and Jerome is a groundhog.

But this cannot be, on Richard's view. Since John only has the RAM determined by 'Jerry is a groundhog and Jerome is a *woodchuck*' in his RS, in order for (7) to be true a correlation would have to map 'groundhog' to 'groundhog' and also to 'woodchuck'.

Granted, in these examples involving belief in conjunctions, sentences ascribing belief in the conjuncts come out true. For example, 'Jane believes that *A* is large' and 'Jane believes that *A* is red' each come out true, provided we are willing to use a separate correlation function for each sentence. Richard might argue that when we utter (5), we "really" have in mind these latter two belief sentences, and so should be satisfied if those latter two sentences come out individually true.

I see two difficulties with this response. First, regardless of what other sentences we have in mind when we utter (5), it seems plausible to claim that (5) itself should come out true. We have strong intuitions about (5) itself, and the substitute offered (namely, the truth of the reports ascribing belief in the conjuncts) doesn't adequately mitigate the implausibility of denying our intuitions about (5). (More cautiously, my claim about (5) is conditional: if our Russellian intuitions are to be respected in the case of (6) (namely, that in some contexts, Jane's acceptance of '*A* is large and *A* is red' is sufficient for the truth of (6)), then they should be respected in the case of (5) as well.)

But the more important difficulty with this response is that it is insufficiently general. Suppose Jane's memory is being tested as well as her vision; after the experiment with the balls in which she correctly identifies ball A as being large and ball B as being red, we wait for a half hour, and then ask her what she remembers about the balls. Her response is: "Not much. I remember that either ball A was large or ball B was red, but nothing more." Now, in an appropriately Russellian context we ought to be able to report her belief by saying "Jane believes that either ball A is large or ball A is red", but this is precluded by Richard's theory, just as the truth of (5) was precluded in the case considered above. And now Richard cannot reply by appealing to the truth of reports about Jane's beliefs in the disjuncts, for none of the following is true:

> Jane believes that ball A is large.
>
> Jane believes that ball A is red.
>
> Either Jane believes that ball A is large, or Jane believes that ball A is red.

6. Yet Another Solution: Correlation Relations

The obvious fix to the problem of section 5 is to change correlation functions to relations. A *correlation relation* is a relation that holds between annotations that have the same "referent part". Hence, a correlation R could hold between ⟨'Twain', Twain⟩ and ⟨'Clemens', Twain⟩, but never between ⟨'Twain', Twain⟩ and ⟨'Updike', Updike⟩. A RAM p represents a RAM q under correlation relation R iff q can be obtained by replacing every annotation a in p by some annotation or other a' such that Raa'. We have Obey(s, R, u) iff for every restriction $\langle u, a, S\rangle \in s$ and every annotation a', if Raa', then $a' \in S$. Call the "relational theory" the theory gotten by making this change to Richard's original restriction-based theory. On its face anyway, the relational theory seems an acceptable patch for the difficulties of the previous section.[21]

[21] Richard would not care for the relational theory. For consider the following argument:

> i) $\exists x \exists y(x = y$ and x is a planet and y is a planet and John believes that he saw x rise, then y rise, then x set, then y set)
>
> ii) therefore, $\exists x \exists y(x = y$ and x is a planet and y is a planet and John believes that he saw x rise, then y rise, then y set, then x set)

Richard regarded it a virtue of his theory that it made this argument invalid. However, the argument is valid on the new theory (assuming as Richard does that no restrictions are allowed on variables). However, I don't think this is conclusive evidence against the relational theory.

7. Combining the Solutions

We have examined two objections to Richard's view. Each time, we considered a modification to his theory to avoid the difficulty. In section 3 we considered changing the restrictions to permissions. In section 6 we considered changing correlation functions to correlation relations. In each case, the modification seemed acceptable. But we must now consider what happens when we *combine* these solutions.

On the combined theory, we have correlation relations as before. We have Permit(s, R, u) iff for every annotations a and a' such that Raa', there is some permission $\langle u, a, S\rangle \in s$ such that $a' \in S$. The trouble comes when we recall the problem in section 3 that (4) and (4′) both turn out true as uttered in Charlie and my context:

(4) Amanda believes that Twain is dead

(4′) Amanda believes that Twain is not dead

This seemed to imply that Amanda is irrational. The reply was that this would imply that Amanda is irrational only if (4) and (4′) are true when evaluated according to a single correlation. If correlations are functions, then no correlation makes both true, since Amanda accepts no pair of sentences of the form:

T is dead

T is not dead

But now that we have allowed correlations to be non-functional relations, this reply is unavailable, for a single correlation relation makes both (4) and (4′) true. A correlation relation that allows 'Twain' to represent both 'Twain' and words like 'he' with which Amanda refers to the man in the picture (i.e. Twain) is allowed by the context's permissions (namely P_1, P_2 and P_3). For the same reason, even the following sentence turns out true:

Amanda believes that Twain is dead and Twain is not dead

Richard says of this argument that it is "far from transparently valid" (Richard, 1990, 153). But I think it is also far from transparently invalid.

for Amanda would utter sincerely 'Twain is dead but *he* is not dead' while pointing at the picture. This is an unfortunate consequence. Amanda, recall, is a leading logician, and would *never* believe a contradiction.[22]

8. Richard and Quantifying In

For a final objection, which is independent of the above, consider the following argument forms:

A1: (i) Twain is a famous author and Odile believes that Twain is dead

(ii) Therefore, $\exists x$(x is a famous author and Odile believes that x is dead)

A2: (i′) Twain is a famous author, and Odile does not believe that Twain is dead

(ii′) Therefore, $\exists x$(x is a famous author, and Odile does not believe that x is dead)

According to Richard, A1 is "undeniably valid".[23] As we'll see below, while his theory implies that A1 is valid, it also has the consequence that A2 is *invalid*. This strikes me as being quite implausible. Each seems valid to me, but what seems even more plausible is the weaker claim that A1 is valid if and only if A2 is valid.

The asymmetry between "positive" and "negative" existential generalization arises for Richard as the result of his analysis of the semantics of "quantifying in". First, he allows RAMs determined by open sentences (relative to an assignment

[22] It might be thought that turning correlation functions into relations ruins Richard's account of the Pierre puzzle all by itself. But Richard can make an alternate response to the Pierre puzzle: when we move from saying "Pierre believes that London is pretty" to "Pierre believes that London is not pretty" we switch contexts; in each of the original contexts, one of the sentences is false, because that context has restrictions that rule out the correlation relations that would make that sentence false. Pierre is not irrational because in such contexts, it is not the case that both ascriptions are true. Granted, there are some contexts in which both turn out true, but this was true on Richard's original theory (see Richard (1990, 180)).

[23] Richard (1990, 152). Since Richard claims to have intuitions about A1, I take it that he assumes the quasi-logical (ii) to have the same truth conditions as some sentence of English, perhaps 'There is some famous author such that Odile believes that s/he is dead'; similarly for (ii′).

to the variables). For example, if variable 'x' is assigned Twain, then the open sentence 'x is dead' determines the RAM $\langle\langle$'is dead', **deadness**$\rangle$, $\langle$'x', Twain$\rangle\rangle$. When it comes to evaluating an open belief sentence, such as 'Odile believes that x is dead', relative to an assignment of Twain to 'x', we need to know what restrictions in the context govern the use of 'x', and here Richard has a special requirement: there can be *no* restrictions on variables short of identity of Russellian content. Thus, when it is assigned Twain, 'x' can represent any term whose Russellian content is Twain—'Clemens', 'Twain', 'I' (when uttered by Twain), 'that guy' (when uttered under appropriate circumstances), etc. Richard makes this requirement in order to insure the validity of arguments like A1. To say that A1 is valid is to say that (ii) is true in any context of utterance in which (i) is true. Now, if (i) is true (in any context), then Odile has some RAM in her RS determined by a sentence of the form $\ulcorner\alpha$ is dead$\urcorner$, where α refers to Twain. Since there are no restrictions on variables, relative to any assignment to the variables that assigns Twain to 'x', a correlation mapping 'x' to α will be permitted in the context, and hence 'Odile believes that x is dead' turns out true under that assignment. So (ii) turns out true as well. A1 is therefore valid.

But suppose Odile rejects the sentence 'Twain is dead', while accepting 'Clemens is dead', as well as accepting every other sentence of the form $\ulcorner\alpha$ is dead$\urcorner$ where α refers to a famous author. In a context with restrictions that require us to use 'Twain' to represent only Odile's uses of 'Twain', (i′) will be true. But (ii′) will be false in that context. Since no restrictions on variables in any context are allowed, the open sentence 'Odile does not believe that x is dead' comes out false when 'x' is assigned Twain, since Odile *does* have in her RS the RAM determined by 'Clemens is dead'. Moreover, whenever 'x' is assigned any other famous author, 'Odile does not believe that x is dead' comes out false, since Odile accepts $\ulcorner\alpha$ is dead$\urcorner$ for every other name α of a famous author. Since there is a possible context in which (i′) is true but (ii′) is false, A2 is invalid.

If an asymmetry between positive and negative existential generalization is objectionable, this means trouble for theories other than Richard's. Let us, somewhat nonstandardly, call a theory "Fregean" if it allows that the following sorts of sentences are consistent (that is, that there is some possible context of utterance in which both of the following are true—in what follows I will often suppress mention of the context):

(8) S believes that …α…

(9) S does not believe that …β…

where α and β are coreferential proper names, demonstratives, or indexicals. (Note that on my usage, Fregean theories needn't imply that names have descriptive contents). No Fregean theory that obeys certain natural assumptions can validate *both* "positive" and "negative" existential generalization on terms inside the scope of 'believes'.[24] Such theories must choose between an uncomfortable asymmetry and rejecting both kinds of existential generalization. The assumptions are:[25]

(AS1) Quantifying in is intelligible, and has an orthodox semantics, in that, for example, $\ulcorner\exists x(S$ believes that x is $F)\urcorner$ is true iff $\ulcorner S$ believes that x is $F\urcorner$ is true under some assignment to 'x'

(AS2) Negations of belief sentences are unambiguous genuine negations, even negations of open belief sentences; thus, relative to any context and assignment to the variables, $\ulcorner{\sim}(S$ believes that x is $F)\urcorner$ is true iff $\ulcorner S$ believes that x is happy$\urcorner$ isn't true

A Fregean theory must allow the possibility of both (10) and (11) being true:

(10) $\alpha =$ the one and only F, and S believes that $\ldots\alpha\ldots$

(11) $\beta =$ the one and only F, and ${\sim}(S$ believes that $\ldots\beta\ldots)$

(where α and β are names, demonstratives, or indexicals). And if both "positive" and "negative" existential generalization are valid, then the following inferences must be valid:

A3: a) $\alpha =$ the one and only F, and S believes that $\ldots\alpha\ldots$
b) Therefore, $\exists x(x =$ the one and only F, and S believes that $\ldots x\ldots)$

A4: a) $\beta =$ the one and only F, and ${\sim}$(S believes that $\ldots\beta\ldots)$
b) Therefore, $\exists x(x =$ the one and only F, and ${\sim}(S$ believes that $\ldots x\ldots))$

[24] Here I thank an anonymous referee for a suggestion.

[25] For an idea of the issues that would be involved in denying (AS2), see Kaplan (1968, section XI); specifically, note the relationship between his (45), (46), and (47). (AS2) could be tinkered with to accommodate truth value gaps.

If (10) is true, then by A3 so is $\ulcorner \exists x(x =$ the one and only F, and S believes that $\ldots x \ldots)\urcorner$; by (AS1) and the nature of the first conjunct of this sentence, there is one and only one assignment to 'x' under which $\ulcorner S$ believes that $\ldots x \ldots\urcorner$ is true. By (AS2), $\ulcorner{\sim}(S$ believes that $\ldots x \ldots)\urcorner$ is not true under that assignment, and so by (AS1), it follows that $\ulcorner \exists x(x =$ the one and only F, and ${\sim}(S$ believes that $\ldots x \ldots))\urcorner$ is not true. But this last sentence follows from (11) by A4. The conclusion, then, is that if (AS1) and (AS2) are true, then the Fregean cannot accept both negative and positive existential generalization.

The Fregean has a choice of evils here. There is always Quining quantification in as being unintelligible, but this skeptical "solution" flies in the face of our linguistic practice. One could accept the intelligibility of quantification in, but reject each inference—existential generalization is invalid on belief sentences and their negations alike. One wonders here why existentially quantified sentences like (10) make sense at all if they never follow from particular instances; after all, one teaches the use of the existential quantifier by pointing out that any instance whatsoever is logically sufficient for the truth of an existential sentence. One might finally attempt to defend the asymmetry in some way. But why would the validity of existential generalization depend on whether or not the sentence in question is a negation?

One might attempt to defend the asymmetry by appeal to belief **of**. Let us return to the case above which showed that Richard's theory invalidates A2—Odile rejects the sentence 'Twain is dead', while accepting 'Clemens is dead', as well as accepting every other sentence of the form $\ulcorner \alpha$ is dead$\urcorner$ where α is a directly referential term referring to a famous author. The argument proceeds as follows. It seems natural to claim that (i′) is true (since Odile rejects 'Twain is dead'), but that 'Odile believes that Clemens is dead' is true (since Odile accepts 'Clemens is dead'). At the very least, there seem to be some contexts of utterance in which these claims will be correct. Since Odile believes that Clemens is dead, Odile believes **of** Clemens that he's dead. But then it follows that Clemens is such that Odile believes that he is dead. From this we infer that the open belief sentence $\ulcorner$Odile believes that x is dead$\urcorner$ is true when Clemens (i.e. Twain) is assigned to 'x'. But since Odile accepts $\ulcorner \alpha$ is dead$\urcorner$ where α refers to any famous author other than Clemens, it follows that (ii′) is false—there is no famous author who Odile *doesn't* believe to be dead. Thus, the notion of belief **of** may be used in rejecting the validity of A2, for it gives us an independent reason for arguing that (ii′) is false in a context in which (i′)

is true.[26]

The problem here is that there is a parallel argument in support of (ii′). The objector reasons as follows using "positive exportation" for belief **of**:

> Odile believes that Clemens is dead,
>
> therefore, Odile believes **of** Clemens that he is dead

But "negative exportation" seems just as plausible to me:

> Odile does not believe that Twain is dead,
>
> therefore, Odile does not believe **of** Twain that he is dead

The objector asserted at the beginning of the argument that Odile doesn't believe that Twain is dead; negative exportation would then imply that Odile doesn't believe **of** Twain that he is dead. But then, by reasoning parallel to the objector's, it follows that Twain is such that Odile doesn't believe that he is dead; hence, ⌜Odile believes that x is dead⌝ is true when Twain (i.e. Clemens) is assigned to 'x'; hence, (ii′) is true, rather than false as the objector had argued. The objector to the validity of A2 uses positive exportation and cannot accept negative exportation, but this asymmetry seems no more plausible than the original asymmetry between A1 and A2 that was being defended.

There are familiar accounts of belief **of**, like those of W.V.O. Quine and David Kaplan, according to which 'Odile believes **of** Twain that he is dead' is true iff there is some description or name of Twain, ϕ, perhaps of a certain restricted sort, such that ⌜Odile believes that ϕ is dead⌝ is true.[27] On such accounts there would indeed be an asymmetry between positive and negative exportation. But are such accounts correct? It seems to me that our natural language intuitions about belief **of** simply do not support any such asymmetry; I suspect that any intuitions to the contrary are tainted because of familiarity with the theories mentioned above. Locutions like 'belief **of**' seldom occur in everyday English, and so the intuitions in question may be more easily drowned out than some.

Let us return to A1 and A2. I have claimed that we have at the very least intuitions that A1 is valid iff A2 is, and also intuitions that each argument form is valid. The former intuitions are incompatible with Richard's theory, and

[26] I thank Mark Richard for helpful comments here.

[27] See Quine (1956); Kaplan (1968).

the latter intuitions are incompatible with any "Fregean" theory—that is, any theory that denies that coreferential names, demonstratives, and indexicals can be substituted within belief contexts *salva veritate*. Thus, the latter intuitions support the Russellianism of Salmon and Soames.

We have here a conflict between what we might call "logical intuitions" and "direct intuitions". Our logical intuitions that both A1 and A2 are valid support Salmon and Soames. "Direct intuitions", which I grant are very strong, support failure of substitution within belief contexts, and thus support Richard and other Fregean theorists. Both sides have means for explaining away the contrary intuitions. As noted at the beginning of the paper, Salmon and Soames explain away the direct intuitions as being merely pragmatic, while opponents can explain away the logical intuitions as resulting from overgeneralizing from more familiar, non-intensional cases. How exactly this stalemate should be broken is not something that I can argue for here; I merely submit that the Fregean denial of our logical intuitions about existential generalization counts against Fregean theories in the final reckoning.

References

Crimmins, Mark (1992). "Context in the Attitudes." *Linguistics and Philosophy* 15: 185–98.

Crimmins, Mark and John Perry (1989). "The Prince and the Phone Booth: Reporting Puzzling Beliefs." *Journal of Philosophy* 86: 685–711.

Grice, H.P. (1975). "Logic and Conversation." In Peter Cole and Jerry Morgan (eds.), *Syntax and Semantics, Volume 3: Speech Acts*, 41–58. New York: Academic Press. Reprinted in Grice 1989: 22–40.

— (1989). *Studies in the Way of Words*. Cambridge, Mass.: Harvard University Press.

Kaplan, David (1968). "Quantifying In." *Synthese* 19: 178–214.

— (1989). "Demonstratives." In Joseph Almog, John Perry and Howard Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press.

Kripke, Saul (1972). "Naming and Necessity." In Donald Davidson and Gilbert Harman (eds.), *Semantics of Natural Language*, 253–355, 763–769. Dordrecht:

Reidel. Revised edition published in 1980 as *Naming and Necessity* (Cambridge, MA: Harvard University Press).

— (1979). "A Puzzle About Belief." In Avishai Margalit (ed.), *Meaning and Use*, 239–83. Dordrecht: Reidel.

Lewis, David (1979). "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8: 339–59. Reprinted in Lewis 1983: 233–249.

— (1983). *Philosophical Papers*, *Volume 1*. Oxford: Oxford University Press.

Quine, W. V. O. (1956). "Quantifiers and Propositional Attitudes." *Journal of Philosophy* 53: 177–187. Reprinted in Quine 1966: 185–196.

— (1966). *The Ways of Paradox*. New York: Random House.

Richard, Mark (1990). *Propositional Attitudes*. New York: Cambridge University Press.

— (1993). "Attitudes in Context." *Linguistics and Philosophy* 16: 123–48.

Salmon, Nathan (1986). *Frege's Puzzle*. Cambridge, Mass.: MIT Press.

Salmon, Nathan and Scott Soames (eds.) (1988). *Propositions and Attitudes*. Oxford: Oxford University Press.

Soames, Scott (1987). "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15: 47–87. Reprinted in Salmon and Soames 1988: 197–239.