

Simple Sentences, Substitutions, and Mistaken Evaluations¹

David Braun

Jennifer Saul

Abstract: Many competent speakers initially judge that (i) is true and (ii) is false, though they know that (iii) is true.

- (i) Superman leaps more tall buildings than Clark Kent.
- (ii) Superman leaps more tall buildings than Superman.
- (iii) Superman is identical with Clark Kent.

Semantic explanations of these intuitions say that (i) and (ii) really can differ in truth-value. Pragmatic explanations deny this, and say that the intuitions are due to misleading implicatures. This paper argues that both explanations are incorrect. (i) and (ii) cannot differ in truth-value, yet the intuitions are not due to implicatures, but rather to mistakes in evaluating (i) and (ii).

1. The Puzzle

We'll begin with some intuitions. First, (1) seems true, while (1*) does not.

- (1) Superman leaps more tall buildings than Clark Kent.
- (1*) Superman leaps more tall buildings than Superman.

Next, a situation in which (2) is true could, seemingly, be one in which (2*) is false.

- (2) Clark Kent went into a phone booth, and Superman came out.
- (2*) Clark Kent went into a phone booth, and Clark Kent came out.

The intuitions we have about these sentences are puzzling, because there are good reasons for thinking that they are correct and also good reasons for thinking that they are *incorrect*.²

But perhaps you did not have these intuitions.³ No matter: it's undeniable that many competent, rational, relevantly well-informed speakers who understand these sentences *do* have these intuitions, at least initially--the reactions of readers to Saul (1997) are sufficient to establish this. The existence of such speakers is all we need to generate the puzzle in which we are interested. In what follows, we shall sometimes call the intuitions of such speakers the *standard* or *typical* intuitions. We will also continue to assume that you had the standard intuitions. If you did not, please assume that we are speaking of the intuitions of some speaker who did.

Are your intuitions about the sentences above correct or incorrect? There are apparently compelling reasons to think that they are *incorrect*. For, obviously, (1*) is false. Also, the following identity sentence is true.

(3) Superman is identical with Clark Kent.

But (1*) seems to follow from (3) and (1). Thus it seems that if (1*) is false, then so is (1), and so your intuition that (1) is true is incorrect. A similar bit of reasoning seems to show that your intuitions about (2) and (2*) are incorrect.⁴

But there are also compelling reasons to think that your intuitions are *correct*. After all, you understood sentences (1) and (1*). You are also (we assume) a competent, rational, and relevantly well-informed speaker. Therefore, your judgments about the truth-values of (1) and (1*) are likely to be correct. Moreover, you knew that (3) is true, and so that Superman is identical with Clark Kent; so if (1*) really did follow from (1)

and (3) by a simple substitution inference, then you would have inferred that (1) is false. Similarly, you would have inferred that (2) and (2*) cannot differ in truth-value. But you didn't.

Thus there is a seemingly strong argument that your intuitions are correct, and a seemingly strong argument that your intuitions are incorrect. So we're left with a puzzle: which argument is unsound, and why? We call this *the puzzle of resistance to substitution in simple sentences*. By 'simple sentence', we mean a sentence that does not contain any quotational, psychological, or other obviously non-extensional contexts. Many other pairs of simple sentences also provoke similar *anti-substitution intuitions*, for example, (4)-(5*).⁵

(4) Clark Kent arrived at the scene of the rescue just after Superman left.

(4*) Clark Kent arrived at the scene of the rescue just after Clark Kent left.

(5) Lex Luthor has hit Superman several times, but has never hit Clark Kent.

(5*) Lex Luthor has hit Superman several times, but has never hit Superman.

In this paper, we offer a solution to this puzzle. According to our solution, your intuitions about (1)-(2*), are incorrect: (1) is false, and (2) and (2*) cannot differ in truth value. But the sorts of errors that we think you make are very different from those which have been previously attributed to speakers in discussions of the above puzzle, and others like it.

2. Background: Simple Sentences, Belief Sentences, and Previous Attempts to Solve the Puzzle

In a previous paper, Saul (1997) presented a selection of simple sentences that provoke anti-substitution intuitions. She argued that they pose a difficulty for theorists who oppose neo-Russellian theories of *belief sentences*, such as the theories of Nathan Salmon (1986, 1989) and Scott Soames (1988, 1995). Salmon and Soames say that pairs of belief sentences that differ only in containing different co-referring names, such as (6) and (7), express the same proposition. They say that standard intuitions that the sentences may differ in truth-value are due to pragmatic factors.

(6) Lois believes that Superman flies.

(7) Lois believes that Clark flies.

Now consider a theorist who rejects Salmon's and Soames's view, on the grounds that it conflicts with typical anti-substitution intuitions about belief sentences. What should such a theorist say about typical speakers' intuitions concerning the *simple* sentences (1)-(2*)? If she says that these simple-sentence intuitions are *correct*, then it seems that she must say that (1) does *not* follow from (1*) and (3). Saul (1997) argued that accounts that allow for this had implausible consequences. But if such a theorist says that typical speakers' anti-substitution intuitions about simple sentences are *incorrect*, then she should say why she trusts typical speakers' anti-substitution intuitions concerning belief sentences more than she trusts their anti-substitution intuitions concerning simple sentences. Furthermore, such a theorist would need to explain away typical intuitions about the simple sentences, and (Saul claimed) would surely hypothesize *pragmatic* differences between the pairs of simple sentences in order to do so. But if this theorist thinks that pragmatics suffices to explain the anti-substitution intuitions in the case of

simple sentences, why does she not think that pragmatics suffices in the case of belief sentences?

Critics of neo-Russellianism have responded to Saul's challenge by taking up both of the preceding options. Graeme Forbes (1997, 1999) and Joseph Moore (1999, 2000) say that the anti-substitution intuitions about simple sentences are *correct*; they say that utterances of (1) and (1*) can semantically express distinct propositions that differ in truth-value, as can (2) and (2*).^6 Alex Barber (2000) maintains that the anti-substitution intuitions for simple sentences are *incorrect*, and attempts to explain away incorrect intuitions by appealing to pragmatics, arguing that his implicature-based account combines naturally with a neo-Fregean semantics that rejects substitution in belief reports.

Saul (1997, 1999, 2000) has criticized Forbes's and Moore's semantic solutions at some length. We think that her criticisms pose serious problems for their views. In this paper, however, we raise additional objections that apply to both semantic solutions (like theirs) and implicature-based solutions (like Barber's). The objections turn on a feature common to both sorts of accounts. The objections also serve to motivate the new solution we propose, which rejects the common feature. On our account, standard intuitions about simple sentences like (1)-(2*) are *incorrect*, but the explanation of these mistaken intuitions does not invoke implicatures. Thus we reject the widespread assumption, formerly held by Saul, that any theory that says that standard intuitions are incorrect must offer a *pragmatic* explanation of your intuitions.^7 According to our alternative solution, you entertained the propositions that (1)-(2*) semantically express when you read (1)-(2*). The propositions that (1) and (1*) semantically express are both

false, and those that (2) and (2*) express cannot differ in truth-value. However, your *evaluations* of those propositions for truth-value, and possible differences in truth-value, were *mistaken*. Thus we call our account ‘the mistaken evaluation account’.

In what follows, we first describe the existing semantic and implicature-based solutions to the puzzle, and present new criticisms of them. We next expose a common assumption of these accounts that we find dubious. We then present our own account, including some explanations of your mistakes in evaluation. We end with some comments on substitution in belief sentences.

3. Semantic and Implicature Solutions

3.1. Semantic Solutions

Forbes (1997, 1999) and Moore (1999, 2000) have offered accounts which are designed to accommodate the typical intuitions about simple sentences in the most straightforward way: by matching them. According to them, there are situations in which an utterance of (2) is true while an utterance of (2*) is false; and there are utterances of (1) that are true. When you read the inscriptions of (1)-(2*), you correctly grasped the propositions that they semantically express, and correctly evaluated those propositions, and that is why you gave the (correct) verdicts that you did.⁸

Forbes and Moore differ over exactly which propositions are expressed by utterances of (1)-(2*). For Forbes, a true utterance of (1) expresses a proposition partly about *personae* (or, more strictly, *modes of personification*). This proposition can be roughly expressed with (1F).

(1F) Superman, so-personified, leaps more tall buildings than Clark Kent, so-personified.

For Moore, a true utterance of (1) expresses a proposition about *aspects* of Superman/Clark, one that can be roughly expressed with (1M).

(1M) Superman/Clark's Superman-aspect leaps more tall buildings than Superman/Clark's Clark-Kent-aspect.

The key difference between these accounts is that for Moore, some utterances of 'Superman' and 'Clark Kent' occurring in utterances of (1) do not co-refer, but instead refer to different aspects of the same individual. For Forbes, all utterances of the names co-refer, but some utterances of (1) express propositions that are also about modes of personification.

Both accounts say that utterances of (1) can express different propositions in different contexts. On Moore's view, this occurs because some utterances of 'Superman' and 'Clark Kent' refer to the individual, and not to aspects of him⁹. On Forbes's view, some utterances of (1) do not express propositions about modes of personification.¹⁰ Forbes and Moore hypothesize this sort of contextual variation in order to account for some of our intuitions. Consider, for example, a conversation between two people who have just learned of Superman's double life, and who are now amazedly working out the consequences of this revelation. One of them utters (8).

(8) Wow, so sometimes Clark Kent wears a cape and leaps tall buildings!

This utterance of (8) seems, intuitively, to be true. But if Moorean or Forbesian propositions about personae or aspects were expressed by all utterances of sentences containing 'Superman' and 'Clark Kent', then no utterance of (8) would be true.¹¹

Moore and Forbes adopt a common explanation of this contextual variation. They distinguish between two sorts of contexts, enlightened and unenlightened. Enlightened contexts are those in which the conversational participants are aware of the relevant double lives, while unenlightened contexts are ones in which the conversational participants are not aware of such facts. In unenlightened contexts, the conversational participants do not know that reference to aspects or modes of personification might be called for, and so utterances of the names refer only to individuals.¹² In unenlightened contexts, then, utterances of (2) and (2*) express propositions that cannot differ in truth-value, and an utterance of (1), like an utterance of (1*), is simply false.¹³ In enlightened contexts, however, conversational participants are in a position to make reference to aspects or modes of personification, and if their focus is on these rather than individuals, the propositions expressed by their utterances will involve aspects or modes of personification. In these contexts, then, (2) and (2*) may express propositions that can differ in truth-value, and (1) may express a true proposition.

Because they disagree about whether utterances of ‘Superman’ and ‘Clark’ co-refer in true utterances of (1), Forbes and Moore would reply differently to the argument that your intuitions are incorrect (which appeared in our introduction). Forbes would reject the assumption that (1*) follows from (1) and (3), while Moore would hold that (3) is false, in the context in which you thought about (1) and (1*).¹⁴

3.2. *Implicature-based Solutions*

We turn now to describing solutions that (i) hold that your intuitions about simple sentences are *incorrect*, and (ii) explain away your incorrect intuitions by appeal to

implicatures. These theories come in (at least) two flavours, neo-Russellian and neo-Fregean.

A neo-Fregean theory, like Barber's, holds that, in every context, (1) and (1*) express distinct, but necessarily equivalent, propositions.¹⁵ Furthermore, identity sentence (3) and identity sentence (3*) express distinct necessary propositions in every context.¹⁶

(3) Superman is identical with Clark Kent.

(3*) Superman is identical with Superman.

Neo-Fregeans have a straightforward explanation for how an *unenlightened* speaker can think that (1) is true and (1*) is false, and that (2) and (2*) can differ in truth-value.

Unenlightened speakers don't believe that Superman is Clark Kent. That is why they fail to substitute, and so have incorrect intuitions about (1)-(2*). However, *enlightened* speakers' incorrect intuitions about (1)-(2*) are *prima facie* puzzling, for these speakers do believe that Superman is Clark Kent, and so are in a position to make the correct substitution inference.

This is the point at which Barber appeals to conversational implicatures. On his view, utterances of (1)-(2*) sometimes conversationally implicate propositions that really do, or can, differ in truth-value. These implicated propositions are very like the propositions ((1F) or (1M)) that Forbes and Moore maintain are semantically expressed by the utterances.¹⁷ When an utterance of (1) conversationally implicates a proposition like (1F) or (1M), an enlightened speaker entertains that implicated proposition; he then consults his relevant beliefs, and correctly judges that this proposition is false. His

judgment that this implicated proposition is false causes him to (incorrectly) think that the utterance of (1) is false. A similar story can be told about (2) and (2*).

In the fourth paragraph of this paper, we presented an argument that your intuitions are correct. Barber can respond to it by pointing out that it implicitly assumes that a rational and relevantly well-informed speaker who understands utterances of the sentences judges that the utterances are true iff he judges that the propositions that they *semantically express* are true. This assumption is false. Even competent, rational, and well-informed speakers sometimes judge that an utterance is true because they judge that the proposition it *implicates* is true. Furthermore, the argument assumes that if (1) followed from (1*) and (3), then you would perform the substitution inference. But that is not so, as you may be confused about what the utterance of (1) semantically expresses, or exclusively focused on what it implicates.

Neo-Russellians can offer a similar implicature-based story. Neo-Russellian semantic theories say that the only contribution a name ever makes to the proposition expressed by an utterance of a sentence containing it is its referent. This means that all utterances of (1) and (1*) express the same proposition, as do all utterances of (2) and (2*), and (3) and (3*). Anyone who believes the proposition expressed by (3), then, believes the proposition expressed by (3*), as they express the same proposition. So (in contrast to neo-Fregeans) neo-Russellians *cannot* say that the *unenlightened* fail to believe the proposition that Superman is Clark Kent—they believe this simply by virtue of believing that Superman is Superman. They cannot, then, explain unenlightened speakers' anti-substitution intuitions by appeal to this ignorance. Instead, neo-Russellians typically say that a single proposition can be believed under different *guises*

or via different *ways of taking* a proposition. An unenlightened speaker believes the proposition that Superman is Clark Kent under a guise corresponding to (3*), but does not believe that proposition under a guise corresponding to (3). Thus he may think that (3*) is true and (3) is false, and may fail to make the relevant substitution inferences in (1)-(2*).¹⁸ However, neo-Russellians cannot explain an *enlightened* speaker's anti-substitution intuitions in this way, for enlightened speakers believe that Superman is Clark Kent under a guise corresponding to (3). Thus enlightened speakers seem to be in a position to recognize that (1) and (1*) have the same truth-value, and similarly for (2) and (2*).

It's at this point that the neo-Russellian theorist appeals to implicatures, in much the way that the neo-Fregean theorist did. The neo-Russellian says that some utterances of (1) implicate a proposition about aspects like those expressed by (1F) or (1M). This implicated proposition is true. Speakers judge that this implicated proposition is true, and so come to believe that the utterance of (1) is itself true.¹⁹ The story about (2) and (2*) is similar. The neo-Russellian response to the argument that your intuitions are correct is much the same as the neo-Fregean's response.

3.3. Critique of semantic and implicature solutions

The semantic and implicature accounts are similar in one respect that is important for our critique: both imply that an utterance of (1) semantically expresses or conversationally implicates a proposition about personae or aspects only if the speaker is thinking about personae or aspects. This, we think, creates problems for both sorts of account.

On Moore's and Forbes's theories, the proposition that (1) semantically expresses varies from context to context. In enlightened contexts, an utterance of (1) *may* semantically express a proposition about aspects, if the conversational participants, including the speaker, are focused on them. (From here on, we use 'aspects' as short for 'aspects or modes of personification'). But in *unenlightened* contexts, an utterance of (1) does *not* express a proposition about aspects; in such contexts, (1) expresses a false proposition entirely about the individual Superman/Clark.

Implicature-based accounts have an analogous consequence. The conversational implicatures of an utterance depend upon the thoughts of the conversational participants. An enlightened speaker may implicate something about aspects in uttering (1), *if* she is focused on aspects when she utters (1). But an *unenlightened* speaker cannot conversationally implicate propositions about aspects. Grice's discussion of conversational implicature clearly supports this conclusion:

[One] who, by (in, when) saying (or making as if to say) that p has implicated that q, may be said to have conversationally implicated that q, provided that (1) he is presumed to be following the conversational maxims, or at least the Cooperative Principle; (2) the supposition that he is aware that, or thinks that, q is required to make his saying or making as if to say p (or doing so in those terms) consistent with this presumption; and (3) the speaker thinks (and would expect the hearer to think that the speaker thinks) that it is within the competence of the hearer to work out, or grasp intuitively, that the supposition mentioned in (2) is required. (Grice 1989: 31.)

Consider condition (3). According to it, a speaker conversationally implicates a proposition about aspects only if the speaker thinks that the hearer can work out that the speaker is thinking about aspects. But if the speaker is *unenlightened*, and not thinking about aspects, then surely the speaker does *not* think that her hearer can work out that the speaker is thinking about aspects. So if the speaker is unenlightened, then her utterances of (1) do not implicate propositions about aspects.²⁰

Thus if either a semantic or implicature account is correct, an utterance of (1) semantically expresses or conversationally implicates a proposition about aspects *only if* (a) the speaker is enlightened and (b) the speaker is thinking about aspects. Moreover, if either account is correct, competent speakers, like you, should know (at least tacitly) that a speaker communicates something about aspects only if the speaker is thinking about aspects.

We can now see that both accounts seem to make certain false predictions about your judgments. If either account were correct, then you would have known that the authors of (1)-(2*) were trying to communicate something about aspects only if they were enlightened and thinking about aspects. So, if these accounts were correct, you would surely have paused to consider whether the authors were thinking about aspects, before passing judgment on the truth-values of (1)-(2*); and you would have withheld judgment if you did not know, or were not sure, whether the authors were focused on aspects. But these predictions are incorrect. You did not pause to consider the knowledge and interests of the authors before making your judgments, and you did not withhold judgment on the sentences. On the contrary, you quickly and confidently made your judgments without any consideration of such matters.²¹

A defender of a semantic or implicature account might reply that, when a typical reader encounters (1)-(2*), he automatically and unconsciously assumes that the authors are enlightened, and are intending to communicate something about aspects of Superman/Clark. This reply may have some initial plausibility when it comes to explaining the intuitions of people who (like you) encounter these sentences in a philosophical article whose title suggests that the authors will discuss substitution puzzles. Such readers might simply assume (incorrectly) that the authors intend to communicate propositions about aspects of Superman/Clark.²² We doubt, however, that people who encounter these sentences outside philosophy journals make such unconscious assumptions; these speakers' lack of hesitation still needs to be explained. Moreover, no such reply can be made to our next objection.

Suppose now that we tell you that sentence (1) was uttered by Lois in a conversation with her friend Myrtle about why Superman is so much more desirable than Lois's dull colleague Clark. Both of them falsely believe that Superman and Clark Kent are distinct individuals. If you are like most speakers, these details will not change the intuitions that you had at the start of this paper.²³ You will still judge that (1) is true.²⁴ But neither the semantic nor the pragmatic accounts we have described can explain this intuition of yours. On these accounts, an *unenlightened* speaker's utterance cannot semantically express, or conversationally implicate, a proposition about aspects. Lois is unenlightened. So, on these accounts, her utterance of (1) can neither semantically express nor conversationally implicate a proposition about aspects. Without such a proposition, said or implicated, none of these accounts can explain your intuition that (1) is true.

3.4. Some Responses and a Misinterpretation

We now wish to consider two responses to our criticisms of the semantic and implicature accounts. We also want to block a misinterpretation of our views that our criticisms could inspire.

Here is the first response: a confused advocate of a semantic or implicature account might claim that your intuition about the Lois-Myrtle case is due to your entertaining an aspect-proposition that Lois's utterance neither expresses nor implicates. But if Lois's utterance neither semantically expresses nor conversationally implicates an aspect-proposition, then your intuition is not explained by either semantics or implicatures. Therefore, this reply is inconsistent with the semantic and implicature accounts of your intuitions.

The second response we wish to consider is more interesting. In "Did Clinton Lie?" (2000), Moore maintains that, in some cases, one proposition is said relative to the audience's context while a different proposition is said relative to the speaker's context. Here's one example that Moore uses to support this theory of relativized expression of propositions: Jack (thinking of Clinton) exclaims "I'm shocked—the president had improper relations!". Jack's audience, Jacques, takes him to be discussing Mitterand. Moore says that, relative to Jacques's context, Jack said something about Mitterand, while relative to his own context Jack said something about Clinton. Our example involves a mismatch between the speaker's and audience's degrees of enlightenment. Moore might therefore maintain that Lois's utterance semantically expressed a proposition about aspects relative to the audience's context, though not relative to hers.

But this hypothetical reply ignores crucial facts about our case. The examples Moore gives to support his view involve situations in which conversational participants are significantly unaware of each other's thoughts and intentions. We can see the importance of this ignorance by extending Moore's example. Suppose that a third-party explains to Jacques that Jack was thinking of Clinton. Jacques will no longer take Jack to be talking about Mitterand, and it is implausible to suppose that Jack said something about Mitterand relative to Jacques's new context. Now in the case we presented, you know that Lois is unenlightened. Thus you know enough about Lois's thoughts and intentions to know that she is not making a claim about aspects. Thus a hypothetical Moorean claim that Lois said something about aspects, relative to your context, is implausible.

Finally, we turn to a potential misinterpretation of our views. Some readers might take us to be claiming that (i) speakers cannot use sentences like (1)-(2*) to implicate propositions about aspects, and (therefore) (ii) speakers' intuitions about utterances of (1)-(2*) can never be explained by implicatures concerning aspects or their ilk. We are not making these claims. Indeed, we think that almost any sentence can be used to implicate almost any proposition, *in the right sort of context*. Thus, *in certain special contexts*, utterances of these sentences can be used to implicate propositions about aspects (or about roles, personae, and similar matters). Suppose, for instance, that Jonathan and Martha Kent know that their adopted son Clark leads a double life, and are discussing his activities. Martha might say to Jonathan, "Clark performs feats when he is occupying his Superman-role that he would never perform when he is not occupying that role. For instance, Superman leaps more tall buildings than Clark". In this context,

Martha's utterance (almost certainly) implicates a proposition about roles, and Jonathan's intuitions about her utterance might be explained by his grasping that proposition. But this context obviously provides a lot of support for such an implicature. This support is absent from the contexts in which you had the standard anti-substitution intuitions about (1)-(2*). We have here argued only that implicatures concerning aspects do not explain the intuitions that you had in those contexts (e.g., when reading the sentences at the beginning of this paper or when considering the Lois-Myrtle example).

The semantic and implicature accounts, then, cannot explain standard initial intuitions about (1)-(2*), and cannot explain the persistence of these intuitions in the Lois-and-Myrtle example. These theories do not, then, solve the puzzle that is the topic of this paper.

4. Our Solution

4.1. A False Principle

The semantic and implicature accounts agree on an important point: when you read (1) and (1*), you entertained some propositions that differ in truth-value, and when you read (2) and (2*), you entertained some propositions that *can* differ in truth-value. The semantic and implicature accounts disagree only about whether the propositions that you entertained were semantically expressed or conversationally implicated. Indeed, nearly all attempts to explain competent, knowledgeable speakers' intuitions about utterances (whether simple or not) hypothesize that these speakers grasp propositions that (i) are either semantically expressed or conversationally implicated by the utterances and (ii) have the (possible) truth-values that the speakers attribute to the utterances. Thus,

one can easily get the impression that nearly all theorists who write about speakers' intuitions accept what we call *The Matching Proposition Principle*.

(MP) *The Matching Proposition Principle*

Suppose that a competent, rational, relevantly well-informed speaker hears and understands an utterance U of a sentence, and judges U to have a (possible) truth-value T. Then there is some proposition P such that:

- (a) U either semantically expresses P or conversationally implicates P; and
- (b) P has (possible) truth-value T.²⁵

No one, to our knowledge, has explicitly endorsed (MP). But the fact is that, when it comes to explaining the truth-conditional intuitions of well-informed speakers, the literature is filled with semantic explanations and implicature-based explanations, and not much else.²⁶ Thus, the literature looks *as if* it were written by authors who tacitly accept (MP). As a result, it becomes very natural for those who read the literature to think that semantics and implicatures are the only options. We think, however, that (MP) is false. The Lois-and-Myrtle example in the last section (if successful) *shows* that it is false, for Lois's utterance neither semantically expresses nor conversationally implicates a proposition that matches your intuitions in the way that (MP) specifies.

In fact, we think it is rather strange that theorists have for so long seemed to adhere to (MP). Consider again what we need to do, if we take certain intuitions about (possible) truth-values to be inaccurate: we need to explain how it is that people make wrong judgments about whether an utterance of a sentence has a certain (possible) truth-value, (given certain facts). *Prima facie*, people could go wrong in any number of ways. To see this, it may be helpful to consider people's errors in two experiments that have

been discussed by cognitive psychologists, the Wason selection task and the Moses illusion.

The Wason selection task involves—in its broadest outline— asking people to judge which facts they need to know in order to evaluate the truth of a conditional statement. Wason’s original task (Wason, 1966) presented subjects with cards that they were told had a letter on one side and a number on the other. They were then asked to say which cards they would need to turn over in order to judge whether the experimenter was speaking truthfully when he said, ‘if a card has a vowel on one side, then it has an even number on the other side’ (Wason 1966, p. 146). People are, in general, very bad at this task. They make terrible decisions as to which information they would need to decide whether the claim in question is true or false.²⁷ There is an enormous literature on this task, and there are many, many explanations of why people make the mistakes that they do. Some do turn on pragmatics, but others involve cheater-detection modules, availability effects, confirmation biases, matching biases, and reasoning schemas—to name just a few.²⁸ The judgments called for in the Wason test are judgments about what information is relevant to a decision regarding truth-value. Such judgments are clearly importantly similar to the sorts of judgments regarding (1)-(2*) that we are trying to explain. Nonetheless, implicature-based explanations are just one sort among many that have been offered for the Wason reasoning error. We think this is a striking fact.

We can best present the Moses illusion by example.²⁹ Please try to answer the following question.

(9) How many animals of each kind did Moses take on the ark?

If you are like most readers, you answered ‘two’. But, of course, that answer is incorrect, for it was Noah, not Moses, who took animals into the ark (according to the Biblical story). What makes your answer puzzling is that you *knew* this fact about Noah (and Moses). Most people similarly tend to judge that sentence (10) is true, even though they “know better”.

(10) Moses took two animals of each kind on the ark.

On the other hand, people tend not to make these mistakes when the name ‘Nixon’ is substituted for ‘Moses’ in either sentence. Experiments strongly suggest that readers correctly understand the relevant sentences (including the name) and that they are not misled by conversational implicatures. If this is so, then every case in which a knowledgeable person judges that (10) is true is a *counterexample* to (MP).³⁰ Nearly all psychologists who have studied this phenomenon agree that the correct explanation involves the fact that most people associate similar features with the names ‘Moses’ and ‘Noah’, for instance, being a Biblical character, receiving messages from God, and performing important deeds involving water (Moses parted the Red Sea). According to one explanation (Reder and Kusbit, 1991), the overlap of features associated with the two names causes readers to make errors when they draw on their memories to answer the (correctly understood) question or to evaluate the truth of the (correctly understood) indicative sentence.

Our brief discussion of the Wason selection task and the Moses illusion should remind us that our intuitions regarding truth values and possible truth values are determined by complex psychological processes. For instance, when you decide whether some utterance of (1) is true or false, given a certain set of facts, you are influenced by (at

least) which facts you take to be relevant and what you take their relevance to be; how well you recall relevant background facts; how long and hard you think about background facts; how well you reason about the impact these facts should have on your judgment; and any number of biases and the like. In principle, errors may occur at any stage. Thus, there is no reason to assume that an error in your judgments about (1)-(2*) could only be due to a confusion of semantics with pragmatics. In particular, there is no reason to assume (as implicature and semantic theorists seem to) that you could *not* have made an error in determining the truth-value (according to your beliefs) of the (implicated or expressed) proposition that you entertained when you read (1). In fact, we think that this is precisely where you made an error. Thus we think that (MP) is false. Below, we describe how you might make such an error.

4.2. *Our Positive Account*

Here is what we think the truth of the matter is, beginning with your intuitions about (1) and (1*). The propositions that (1) and (1*) semantically express are propositions about the individual Superman/Clark, and not propositions about aspects. When you read (1)-(1*), you entertained these propositions and you did *not* entertain any propositions about aspects. But you made some mistakes when you evaluated these semantically expressed propositions for truth. For instance, according to your beliefs, the proposition that (1) semantically expressed was false, but you came to believe it was true. Thus you came to think that the relevant inscription of (1) was true, even though the proposition that you entertained as a result of reading (1) was false. Finally, you did not consider the proposition that Superman is Clark Kent during this evaluation procedure; or

if you did, you failed to go through the sort of reasoning that would have allowed you to detect your mistake. We shall call this the *Basic Version of the Mistaken Evaluation Explanation* of your simple sentence intuitions, or the ‘Basic Explanation’, for short.

Our discussion of (MP), and Wason and Moses, should make it clear that the Basic Explanation might correctly explain your intuitions, despite its inconsistency with (MP). The Basic Explanation, however, does not describe *how* or *why* you made a mistake in your initial evaluation of the proposition that you entertained—it’s unlikely, then, to be fully satisfying without a bit of supplementation. We suspect that more detailed explanations of the initial mistake may vary from individual to individual, and may vary across different occasions for the same individual. Furthermore, we think that more detailed explanations of these mistakes can only be discovered empirically. Nevertheless, we shall engage below in some speculations as to the reasons for your mistaken evaluation.

We suspect that, because Superman/Clark leads a double life, you maintain two cognitively separated sets of beliefs about him, one of which is associated with the name ‘Superman’, the other of which is associated with the name ‘Clark Kent’. You do this even though you know that these are names of the same individual. We shall call these two sets of beliefs your ‘two pools of information’. These two pools of information attribute different properties to Superman/Clark. For instance, the ‘Superman’ pool attributes to him the property of leaping tall buildings, while the ‘Clark’ pool does not.³¹ You also associate different images with the two names. When you *quickly* evaluated the proposition semantically expressed by (1), your pools of information and images appeared to you to support that proposition’s truth. Therefore, you judged that (1) was

true. You did not pause to consider the identity long enough to notice its logical consequences. Or, if you did, you erred in not considering this good reason to alter your original judgment. Let's call this the *Two Pools Version of the Mistaken Evaluation Explanation*, or the 'Two Pools Explanation' for short. Notice that it is just an elaboration on the Basic Explanation. In section 4.8 below, we describe how both neo-Russellians and neo-Fregeans could accept it.

Even the Two Pools Explanation is rather sketchy about how you made your mistake. But to provide a more detailed explanation, we must make further, even more speculative, assumptions about your psychological processes. We think it will be useful to consider such speculative explanations, for two reasons. First, the forthcoming speculative explanations suggest that it is *possible* for a person with the right sort of psychological structure (whether or not his psychological structure is exactly like yours) to entertain the propositions that (1)-(2*) semantically express, without entertaining propositions about aspects, and yet still make the same erroneous judgments that you did. Second, we think that consideration of these speculative explanations makes it evident that our Basic and Two Pools Explanations are compatible with a very wide range of assumptions about our psychological makeups.³²

4.3. *Some Cognitive Architecture*

To present our most detailed and speculative explanations of your mistakes, we must (for the moment) make some assumptions about your *cognitive architecture*. Let's assume that humans have two sorts of *mental representation*: sentences in a language of thought and images. The former have structures similar to those of natural language

sentences. For convenience, we shall assume that your language of thought is English.³³ Images, on the other hand, have non-linguistic structures. They represent objects and events in some way similar to maps or photographs or movies.

We'll assume that a person believes a proposition if (and maybe only if) he has a mental sentence in his head that functions in the appropriate belief-like way. Whenever a person has a sentence in his head that functions in the belief-like way, we will say that the sentence is in his *belief box*.³⁴ Similarly, a person entertains a proposition if (and perhaps only if) he has a mental sentence in his head that functions in the appropriate entertainment-like way (different from the belief-like way). Whenever a person has a sentence in his head which functions in this way, we will say that he has that sentence in his *entertainment box*.³⁵ We will also assume that every subject maintains at least one *file* for each person about whom that subject has beliefs. These files are collections of mental sentences containing one of the person's names.³⁶ These sentences are causally related to each other in a particularly intimate way. For instance, if a person consults a file, then the sentences in it become more "active", though some may become more active than others. Consequently, it becomes easier for those sentences to enter that person's entertainment box; moreover, the more active sentences become easier to entertain than the less active ones.

We all know about Superman's double life.³⁷ Thus all of us have the sentence 'Superman is Clark' somewhere in our belief boxes. Nevertheless, Superman/Clark's double life gives us good reason to maintain two distinct files on him, one containing 'Superman' sentences and one containing 'Clark' sentences.³⁸ Files of the former sort contain sentences like 'Superman wears a red cape' and 'Superman fights for truth,

justice, and the American way'. Files of the latter sort contain 'Clark Kent wears glasses' and 'Clark Kent works for the *Daily Planet*'. A given person's 'Superman' file may attribute properties to Superman/Clark that her 'Clark' file fails to attribute, e.g., her 'Superman' file may contain 'Superman flies' while her 'Clark' file may not contain 'Clark flies'. A typical speaker may routinely make additions and subtractions to one file without making the corresponding additions and subtractions to the other. For instance, if she reads 'Superman saved a person who fell off a cliff' in a reliable newspaper, she deposits this sentence in her 'Superman' file, but not her 'Clark' file. Moreover, she does *not* (typically) add the corresponding 'Clark' sentence to her 'Clark' file.³⁹

In short, we do not ordinarily perform all of the substitution inferences that are allowed by the sentences in our belief boxes. This failure to perform substitution inferences, and this kind of segregation of 'Superman' and 'Clark' sentences, makes cognitive sense. It takes cognitive effort, and other cognitive resources, to duplicate information from one file to the other (by making substitution inferences), yet because Superman/Clark leads a double life, this duplication often serves no useful purpose. Thus, we often treat 'Superman' and 'Clark' sentences as if they were about different people.

We also associate different images, and imaging routines, with the names 'Superman' and 'Clark'. If asked 'Does Superman wear red boots?', we form an image of a man wearing a cape, not an image of a man in glasses and a business suit. When we read in a newspaper the sentence 'Superman saved a person who fell off a cliff', we generate an image of a man with a red cape, flying and catching a person in mid-air, and not an image of a man in glasses catching a person. One reason that it's reasonable for us

to associate different sorts of image with the two names is that the two sorts of image differ in their *accuracy conditions*. For example, the former image is accurate only if a man, while wearing a red cape, caught a person in mid-air. This image is likely to be accurate. The second image is accurate only if a man, while wearing glasses and a business suit, caught a person. It is likely to be *inaccurate*.⁴⁰ If we were just as likely to form a ‘glasses’ image as a ‘caped’ image when we heard a ‘Superman’ sentence, many more of our images would be inaccurate.

4.4. *Story One*

Using these assumptions, we can construct one reasonable explanation (others will follow) of how you came to believe that (1) is true and (1*) is false.

(1) Superman leaps more tall buildings than Clark Kent.

(1*) Superman leaps more tall buildings than Superman.

When you read (1), you entertained the proposition that it semantically expresses, by having sentence (1) in your entertainment box. This is a proposition about individuals and not about aspects. You then tried to evaluate that proposition for truth. Prior to reading these sentences, you had a sentence like ‘Superman leaps tall buildings in a single bound’ in your ‘Superman’ file. You had no such sentence in your ‘Clark’ file. In fact, in your ‘Clark’ file you had sentences such as ‘Clark is a mild-mannered reporter’, which strongly suggest that Clark doesn’t leap any tall buildings. Thus, when you turned to your ‘Superman’ and ‘Clark’ files in order to evaluate the proposition expressed by (1), you found sentences in those files that seemed to support its truth. You didn’t stop to consider the identity sentence ‘Superman is Clark’, or its logical consequences, even

though you had it in your belief box. You failed to do this simply because you (quite reasonably) don't usually do so when you entertain sentences containing 'Superman' or 'Clark'. So you concluded that (1) is true.

When you read (1*), you entertained the proposition that (1*) semantically expresses, by having (1*) in your entertainment box. You thought that (1*) could not be true (perhaps because of its form), so you judged it to be false. Thus you came to believe that (1) is true and (1*) is false.

If you had entertained (3), and considered its logical relations with (1) and (1*), you might have realized that you made a mistake. But, for the reasons discussed above, you didn't. Thus you did not perform the inferences that would allow you to realize that if (1*) is false then (1) must be false also. That is perhaps why you initially judged that (1) is true and (1*) is false, despite the fact that you believed (in a suitable way) that Superman is Clark.

To determine whether it's possible for (2) and (2*) to differ in truth-value, you might have placed those sentences in your entertainment box, and tried to determine whether the one can be (syntactically) derived from the other, together with various truths in your belief box. Or you might have entertained (2) and the negation of (2*), to see whether there is any inconsistency between them and sentences in your belief box. If you did this rather hastily, without considering the identity, you might not have realized that there was an inconsistency.

4.5. *Story Two*

An alternative psychological explanation might appeal more heavily to your imaging processes. When you entertained (1), you generated an image of a caped man leaping a tall building and an image of a bespectacled man leaping a tall building. These are images that you might form when you entertain ‘Superman leaps tall buildings’ and ‘Clark Kent leaps tall buildings’. You came to believe that the cape-image accurately represents some events that actually occurred, whereas the glasses-image does not. This led you to think that ‘Superman leaps tall buildings’ is true, whereas ‘Clark Kent leaps tall buildings’ is false, which led you to judge that (1) is true. Again, you did not consider the fact that Superman is Clark, which might have given you pause.

An imagistic explanation seems particularly appealing when it comes to explaining your intuitions about (2) and (2*).

(2) Clark Kent went into a phone booth, and then Superman came out.

(2*) Clark Kent went into a phone booth, and then Clark Kent came out.

When you considered whether (2) could be true while (2*) is false, you tried to form an image of an event that is accurately described by (2) and also an image of an event accurately described by (2*). Considering (2), you formed an image of a bespectacled man going into a phone booth and a caped man emerging. Considering (2*), you formed an image of a bespectacled man going into a phone booth and a bespectacled man emerging. Clearly, there can be a sequence of events that is accurately represented by the first image, but not accurately represented by the second. Thus you might have concluded that it’s possible for (2) to be true and (2*) to be false.

4.6. *Entertaining the Identity*

The above explanations assume that you did not entertain the proposition that Superman is Clark Kent (by having (3) in your entertainment box) as you tried to judge the truth-values of (1) and (1*) and possible differences in truth-value of (2) and (2*). But we think it's also possible that you entertained (3) and nevertheless judged that (1) is true and (1*) is false, and that (2) and (2*) could differ in truth-value. For you might have gone through one of the evaluation procedures we described above, and thus come to be confident that (1) is true and (1*) is false. You may then (or at the same time) have come to entertain (3). But you may not have considered the consequences of combining (1) and (1*) with the identity, simply because you were already confident of your answer. Similar points could hold for (2) and (2*).

We do not think that this would have been irrational on your part. We simply cannot take the time to draw out many of the logical consequences of the propositions we believe and entertain before judging whether an English sentence is true. Since in most other cases you quite reasonably do not make the identity substitutions, you quite reasonably failed to do so in this case.

We have now presented several rather detailed, but speculative, possible explanations of your intuitions about (1)-(2*). We shall call these our *Speculative Explanations*. Notice that all of them are consistent with, and are elaborations on, our Basic Explanation and Two Pools Explanation.

4.7. *Our Speculative Explanations, Neo-Russellianism, and Neo-Fregeanism*

Our Speculative Explanations were couched in terms of mental sentences and images. We made very few assumptions about the propositions that people believe and entertain. For example, when we supposed that you had (1) and (1*) in your entertainment box, we assumed that these sentences expressed propositions about the individual Superman/Clark (and not aspects), but we made no assumptions about whether you were thereby entertaining two distinct propositions or a single proposition “twice over”. Similarly, we made no assumption about whether a person who has ‘Superman is identical with Superman’ in his belief box believes the same proposition as someone who has ‘Superman is identical with Clark’ in his belief box.

Neo-Russellians, like ourselves, hold that (1) and (1*) semantically express the same proposition; similarly for ‘Superman is Superman’ and ‘Superman is Clark Kent’. Neo-Russellians can say that a person who has both in her entertainment box is entertaining the same proposition *in two different ways*. Similarly, a person who has (1*) and the negation of (1) in her belief box thereby believes a proposition and its negation, but does so in different ways. Neo-Fregeans, however, hold that (1) and (1*) express distinct propositions, as do the identity sentences. They could maintain that someone who has both (1) and (1*) in her belief box (or entertainment box) thereby believes (entertains) two distinct propositions.⁴¹ Thus both neo-Russellians and neo-Fregeans can accept our Speculative Explanations.

4.8. *Getting Along with Weaker Psychological Assumptions*

Our Speculative Explanations relied on a rather crude mental sentence theory of belief and entertainment, and a crude theory of images. You may have doubts about these theories (as do we, to varying degrees and for varying reasons). But this does not mean that you should reject our Basic Explanation, or even our Two Pools Explanation, for neither depends on psychological assumptions as strong as those of the Speculative Explanations. Below we re-present our Two Pools Explanation in slightly more detail than we did before. Most importantly, we show that both neo-Russellians and neo-Fregeans can accept the Two Pools Explanation.

A neo-Russellian who does not want to make detailed commitments about cognitive processing could say something like the following. (1) and (1*) semantically express the same proposition (in all contexts). You understood (1) and (1*), and entertained the single proposition that they express, but you did so “twice over”: that is, you entertained that single proposition *in two distinct ways*. You tried to judge whether that proposition is true, taken in two distinct ways. You held various beliefs about Superman/Clark in various ways, and these contributed to your judging that certain propositions, entertained in certain ways, were true or false. For instance, you may have believed, *in a ‘Superman’ way* (but not a ‘Clark’ way) that Superman sometimes jumps tall buildings. You may have believed, *in a ‘Clark’ way* (but not a ‘Superman’ way) that Clark is a mild-mannered reporter. Together, these led you to believe the proposition expressed by (1), in a ‘Superman/Clark’ way. You also believed the proposition that Superman is Clark Kent “twice over”, in a ‘Superman/Superman’ way and a ‘Superman/Clark’ way. But, as usual, you (quite reasonably) did not entertain the

identity proposition at all (in any way); or if you did entertain it in the right way, you failed to make the correct inferences in the right ways. A similar story can be told about (2) and (2*).

A neo-Fregean can give a parallel story, but need not mention ways of believing. On the neo-Fregean view, (1) and (1*) semantically express different propositions. You understood (1) and (1*), and entertained the propositions that they semantically express. You believed various Superman-propositions, and various Clark-propositions; for instance, you believed that Superman sometimes jumps tall buildings and that Clark is a mild-mannered reporter. Your beliefs led you to make some initial mistakes in evaluating the proposition semantically expressed by (1). Furthermore, you either failed to entertain the proposition that Superman is Clark, or if you did entertain it, you failed to draw the logical consequences of that proposition together with (1) and (1*). Either way, you did not correct your initial mistake.

Neo-Russellians and neo-Fregeans can respond similarly to the argument in the fourth paragraph of our introduction that concludes that your intuitions were correct. This argument implicitly assumes that you correctly evaluated the propositions semantically expressed by the sentences, according to your beliefs. It wrongly assumes that you made no mistake in evaluation, and that you considered the relevant identity and would have used it to rectify any initial mistakes.

We hope that it is clear that our Basic Explanation and Two Pools Explanation do not rely on claims about mental sentences, files, belief boxes, and imaging routines. Our Basic Explanation assumes only that you entertained the propositions semantically expressed by (1)-(2*), which are propositions entirely about an individual, and that you

made mistakes in evaluating them. Our more detailed Two Pools Explanation assumes that—even when you know the truth about a double life—you sometimes segregate your information about a single individual into distinct “pools”; because you do this, and because you do not regularly draw out all of the logical consequences of propositions that you believe, you sometimes give inaccurate verdicts about the truth-values (and possible differences in truth-value) of utterances of sentences concerning individuals with double lives. We therefore think that any reasonable theory of cognitive processes will give us all that we need to tell our Two Pools story, or at least our Basic story, about why your intuitions are mistaken.

In the rest of this section, we (i) consider subjects whose anti-substitution intuitions are particularly stubborn, and (ii) consider an objection to our explanations.

4.9. Stubborn Ordinary Speakers

We suspect that most ordinary speakers will eventually decide that (1) and (1*) are both false, if they are coached and led step-by-step through some reasoning that shows that they are. Similarly, they will eventually decide that (2) and (2*) cannot differ in truth-value, after coaching. This is obviously compatible with our explanations of the initial judgments. In fact, we take it to fit rather nicely with our hypothesis that the initial judgments are partly due to failures to make certain inferences. The speakers’ understanding of the sentences does not change as they go through the coaching and reasoning. Rather, they just think more carefully about the implications of (1), (1*), and (3).⁴²

Some ordinary speakers, however, might refuse to alter their initial judgments. They might persist in thinking that (1) is true and (1*) is false. They might try to justify their persisting judgments by appeal to something like aspects. They might say that Superman, when dressed in a cape, does leap more tall buildings than Clark, when dressed in a business suit, or claim that Superman, when playing Superman, does leap more tall buildings than Clark, when playing Clark (or when disguised as Clark). They might even say that they meant something like this when they said that (1) is true and (1*) is false.

Do such stubborn ordinary speakers lend any support to the semantic or implicature explanations? We think not. These stubborn ordinary speakers claim that they “meant” something about aspects when *they* uttered ‘(1) is true and (1*) is false’. So they are making claims about *their own* utterances. Thus even if their claims are correct, they do not show that the inscriptions of (1) and (1*) *that they read* semantically expressed or conversationally implicated propositions about aspects. Furthermore, such after-the-fact claims about what one “meant” by past utterances are often unreliable (as are many after-the-fact judgments about one’s states of mind). So, we should not take for granted that these speakers’ claims about their utterances are correct.

4.10. *Stubborn Sophisticated Theorizers*

Sophisticated theorists can make initial mistakes in evaluation very similar to those of ordinary speakers. Suppose that we ask a world-class logician (call her ‘Jo Forbs’, or ‘JF’ for short) whether sentences (1) and (1*) differ in truth-value. She may initially execute the same procedures for answering the question that an ordinary speaker

would. She might consult her ‘Superman’ and ‘Clark’ files, and/or form images, and come to conclude that (1) is true and (1*) is false. She may do this before actively entertaining the identity. On our account, her intuitions are incorrect, but reasonable.

Suppose, now, that we remind JF that ‘Superman is Clark Kent’ is true; suppose that she then starts actively entertaining the identity. Still, she might not draw the correct conclusion that (1) is false (just like (1*)). For JF has already rather confidently judged that (1) is true, and may be under the impression that she previously considered all relevant facts, such as the identity, when she formed this judgment. Moreover, JF, being a logician, is well aware that English does not always work in the way that formal languages do. JF knows, for instance, that the sentences of English are ambiguous and context-sensitive, and so she may conclude that (1) and (1*) are ambiguous or context-sensitive. So she may conclude that, in some contexts, (3) is false (as does Moore), or that, in some contexts, (1*) and (3) are true, while (1) is false (as does Forbes). Thus JF might mistakenly be led to propose an unusual semantic theory for these sentences, and for (2) and (2*), as well.

4.11. An Objection and a Reply

We have argued that your reasonable tendency to maintain separate pools of information about those with ‘double lives’ means that you don’t automatically make all the substitution inferences that you could, and that you may sometimes be very reluctant to do so, even upon careful consideration. But this theory might seem to predict that you would initially resist substitution in *all* simple sentences that mention people or objects with ‘double lives’. And this might seem to be an implausible prediction. After all, there

is a vast literature in the philosophy of language that assumes that substitution inferences are obviously acceptable in sentences like (11) and (11*).

(11) Hesperus appears in the evening.

(11*) Phosphorus appears in the evening.

This literature takes for granted that substitution fails (if it does) *only* in attitudinal and quotational contexts. The reason, perhaps, is that simple sentences that provoke anti-substitution intuitions are hard to come by. So, if our explanations predict that all simple sentences provoke anti-substitution intuitions, then (one might conclude) our explanations are incorrect.

Our explanations do not, however, predict that all substitution inferences will be difficult. First, the Basic Explanation says only that you made a mistake when you evaluated (1)-(2*). It does not commit itself to any details about *how* you made that mistake, and so makes no real predictions about other simple sentences. It's compatible with the Basic Explanation that you find other substitution inferences easy and obvious. Next, the Two Pools and Speculative Explanations are not meant to describe every element of what takes place when you consider simple sentences mentioning those with double lives. Thus all of our explanations may be consistent with more elaborate theories that also explain why some substitution inferences are difficult and some are easy. We'll offer some suggestions for these shortly.

We also want to note that the data invoked in this objection may not be quite so robust as they appear. In most philosophical discussions, audiences are asked to consider the relevant predicative sentences (e.g., (11) and (11*)) either immediately before or immediately after they consider the relevant identity sentence ('Hesperus is

Phosphorus'). We suspect that this plays an important role in producing standard intuitions. In addition, we think that ordinary speakers, unaccustomed to the distinction between intensional and non-intensional contexts, and unaccustomed to discussions of substitution puzzles, are often more reluctant to make the standard substitutions than philosophers. (Some readers may be able to recall teaching experiences that support this claim.)

We do think, however, that there are clear differences in the reactions provoked by sentences like (1) and (1*) and sentences like (11) and (11*). We also think we can offer some explanation of this. There is an obvious difference in overt syntax between our examples and more standard philosophical examples: each of our sample sentences contains two name occurrences, whereas the more usual philosophical sentences contain only one name occurrence. In particular, sentences (1) and (2) contain occurrences of two *different* names. We strongly suspect that this feature of our examples is crucial to explaining why they provoke stronger anti-substitution intuitions than do (11) and (11*).⁴³ We're not sure exactly *why* this feature causes readers to have such strong intuitions. We suspect, however, that the occurrence of two different names in a single sentence creates an unconscious, overridable presumption that the names do not co-refer. Why should this be? One reason might be that speakers who thought that they were discussing just one individual could easily utter a different sentence that made this clear. For instance, instead of uttering (2), a speaker might utter (2a) or (2b).

(2) Clark Kent went into a phone booth, and then Superman came out.

(2a) Clark Kent went into a phone booth, and then he came out.

(2b) Clark Kent went into a phone booth, and then came out.

Consequently, when you read a sentence like (2), you are led to form an especially strong (unconscious) presumption of non-coreference. This increases your reluctance to fully consider all the logical consequences of the identity sentence ‘Superman is Clark Kent’ when considering the truth-values of sentences like (2).⁴⁴ We think that this account might be fleshed out by invoking resources from either Levinson’s work on pragmatics and anaphora, or Discourse Representation Theory, but we won’t attempt to do this here.⁴⁵ In any case, our Basic, Two Pools, and Speculative Explanations could be consistently supplemented with such accounts. The resulting explanations would still attribute mistaken evaluations to typical speakers.

Obviously, we have not provided a complete account of why (1)-(2*) provoke stronger anti-substitution intuitions than most simple sentences. But it is important to note that our competitors are in the same position that we are. Forbes and Moore have postulated a potential for contextual variation in what is said by *all* simple sentences. They need to explain why this potential hasn’t been noticed before, and, in particular, why the assertion that (11) and (11*) have the same truth-value has gone unquestioned for so long. Barber, similarly, needs to explain why the substitution-blocking implicatures which arise with sentences like (1)-(2*) do not arise with most other simple sentences. Clearly, then, the problem we’ve discussed here is not peculiar to our view.

5. Belief Sentences

We shall not argue that our explanation of substitution-resistance in simple sentences should be extended to belief sentences. Indeed, the explanation of substitution-

resistance in belief sentences that we favor (Braun 1998) is importantly different from our explanation of simple-sentence intuitions.⁴⁶ However, we think that the arguments of this paper are relevant to belief sentences in two ways: (i) they cast doubt on certain arguments for and against neo-Russellian theories of belief sentences, and (ii) they suggest a new line of inquiry regarding belief sentences.

Consider the following argument *against* neo-Russellian theories of belief sentences: *All* anti-substitution intuitions of linguistically competent, rational, reflective speakers are correct, and must be explained by semantics. Therefore, such speakers' anti-substitution intuitions about *belief sentences* are correct and must be explained by semantics. Therefore, neo-Russellian theories of belief sentences are incorrect. We suspect that, before reading this paper, some philosophers would have found this argument rather plausible. But we have given reasons to think that anti-substitution intuitions about certain *simple* sentences cannot be correctly explained by semantics. Therefore, we have provided reasons to think that the first premise of this argument is false.

Consider the following argument *for* the neo-Russellian theories of Salmon and Soames.⁴⁷ The intuitions about (1)-(2*) are not correctly explained by semantics. If this is so, then they are correctly explained by implicatures. But if they are correctly explained by implicatures, then (it's likely that) belief-sentence intuitions are also correctly explained by implicatures. Thus, (it's likely that) Salmon and Soames's neo-Russellian theory of belief sentences is correct. Our discussion here gives us reason to reject this pro-Salmon-and-Soames argument. Although we have argued that the simple-sentence intuitions are not correctly explained by semantics, we have also argued that they are not correctly explained by implicatures, either. Therefore, we have given

reasons to think that the second premise of this pro-Salmon-and-Soames argument is false.

Thus, our discussion here provides us with reasons to reject *both* the above argument *against* neo-Russellian theories of belief sentences, *and* the above argument *for* Salmon and Soames's implicature-based explanation of anti-substitution intuitions about belief sentences.

We think that our discussion of simple sentences suggests a new avenue of inquiry regarding belief sentences. Very few theorists have proposed theories of *belief sentences* that conflict with (MP).⁴⁸ Most theorists have tried to explain anti-substitution intuitions about belief sentences by appealing to some proposition, either semantically expressed or conversationally implicated, that matches typical intuitions. These theorists have adhered to (MP), when it comes to explaining belief-sentence intuitions. In our opinion, they have failed to find such intuition-matching propositions (for some criticisms of these theories, see Saul 1992, 1998, 1999a, 1999b). But we have argued that (MP) is false, on grounds independent of belief sentences. Hence, our discussion of simple sentences gives us reason to think that no theory of belief sentences should be rejected *simply* because it is inconsistent with (MP).⁴⁹ Therefore, given the problems with theories that are consistent with (MP), we think that theorists should consider exploring theories of belief sentences that are inconsistent with (MP). We suspect that the correct explanation of belief-sentence intuitions may well be significantly unlike our explanation of simple-sentence intuitions. Nevertheless, we conjecture that the two explanations will be alike in one important respect, namely, in being inconsistent with (MP).⁵⁰

Bibliography

- Barber, Alex. 2000. "A Pragmatic Treatment of Simple Sentences." *Analysis* 60, pp. 300-8.
- Braun, David. 1998. "Understanding Belief Reports." *Philosophical Review* 107, pp. 555-595.
- Braun, David. 2002. "Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions." *Philosophical Studies* 108, pp. 65-81.
- Erickson, Thomas D. and Mattson, Mark E. 1981. "From Words to Meaning: A Semantic Illusion." *Journal of Verbal Learning and Verbal Behavior* 20, pp. 540-551.
- Evans, Jonathan St. B. T., Newstead, Stephen E., and Byrne, Ruth M. J. 1993. *Human Reasoning: The Psychology of Deduction*. Hove, UK: Lawrence Erlbaum.
- Forbes, Graeme. 1997. "How Much Substitutivity?" *Analysis* 57, pp. 109-113.
- Forbes, Graeme. 1999. "Enlightened Semantics for Simple Sentences." *Analysis* 59, pp. 86-91.
- Hannon, Brenda and Daneman, Meredyth. 2001. "Susceptibility to Semantic Illusions: An Individual-differences Perspective." *Memory & Cognition* 29, pp. 449-61.
- Kamas, Eleen N., Reder, Lynne M., and Ayers, Michael S. 1996. "Partial Matching in the Moses Illusion: Response Bias Not Sensitivity." *Memory & Cognition* 24, pp. 687-699.
- Kamp, Hans and Reyle, Uwe. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.

- Kaplan, David. 1989. "Demonstratives." In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes from Kaplan*, pp. 481-564. Oxford: Oxford University Press.
- Kripke, Saul. 1979. "A Puzzle About Belief." In A. Margalit (ed.), *Meaning and Use*, pp. 239-83. Dordrecht: Reidel.
- Levinson, Stephen. 2000. *Presumptive Meanings: the Theory of Generalized Conversational Implicature*. Cambridge, MA: MIT Press.
- Moore, Joseph. 1999. "Saving Substitutivity in Simple Sentences." *Analysis* 1999, pp. 91-105.
- Moore, Joseph. 2000. "Did Clinton Lie?" *Analysis* 60, pp. 250-4.
- Ostertag, Gary. 1998. "The New Frege Puzzle." Paper Presented at the Meeting of the Pacific Division of the American Philosophical Association.
- Perry, John. 1980. "A Problem About Continued Belief." *Pacific Philosophical Quarterly* 61, pp. 317-32. Reprinted in John Perry, 1993, *The Problem of the Essential Indexical*. Oxford: Oxford University Press, pp. 69-90.
- Perry, John. 1997. "Indexicals and Demonstratives." In Bob Hale and Crispin Wright (Eds.), *A Companion to the Philosophy of Language*, pp. 586-612. Oxford: Blackwell.
- Pitt, David. 2001. "Alter Egos and Their Names." *Journal of Philosophy* 98, pp. 531-52.
- Reder, Lynne M. and Kusbit, Gail. 1991. "Locus of the Moses Illusion: Imperfect Encoding, Retrieval, or Match?" *Journal of Memory and Language* 30, pp. 385-406.
- Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.

- Salmon, Nathan. 1989. "Illogical Belief." *Philosophical Perspectives* 3, pp. 243-85.
- Salmon, Nathan and Soames, Scott (Eds.). 1988. *Propositions and Attitudes*. Oxford: Oxford University Press.
- Saul, Jennifer. 1997a. "Substitution and Simple Sentences." *Analysis* 57, pp. 102-8.
- Saul, Jennifer. 1997b. "Reply to Forbes." *Analysis* 57, pp. 114-8.
- Saul, Jennifer. 1998. "The Pragmatics of Attitude Ascription." *Philosophical Studies* 92, pp. 363-98.
- Saul, Jennifer. 1999a. "The Best of Intentions: Ignorance, Idiosyncrasy, and Belief Reporting." *Canadian Journal of Philosophy* 19, pp. 29-48.
- Saul, Jennifer. 1999b. "The Road to Hell: Intentions and Propositional Attitude Ascription." *Mind & Language* 14, pp. 356-75.
- Saul, Jennifer. 1999c. "Substitution, Simple Sentences, and Sex Scandals." *Analysis* 59, pp. 106-112.
- Saul, Jennifer. 2000. "Did Clinton Say Something False?" *Analysis* 60, pp. 255-7.
- Saul, Jennifer. 2001. "Critical Study of Davis, *Conversational Implicature*." *Noûs* 35, pp. 630-641.
- Saul, Jennifer. 2002. "Speaker Meaning, What is Said, and What is Implicated." *Noûs* 36, pp. 228-248.
- Schiffer, Stephen. 1981. "Truth and the Theory of Content." In H. Parret and J. Bouveresse (Eds.), *Meaning and Understanding*, pp. 204-22. New York: de Gruyter.
- Soames, Scott. 1988. "Direct Reference, Propositional Attitudes, and Semantic Content." In Salmon and Soames 1988, pp. 197-239. Originally appeared in *Philosophical Topics* 15 (1987), pp. 47-87.

Soames, Scott. 1995. "Beyond Singular Propositions?" *Canadian Journal of Philosophy* 25, pp. 515-50.

Soames, Scott. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.

Wason, P. 1966. "Reasoning." In B. M. Foss (Ed.), *New Horizons in Psychology*, pp. 135-151. Harmondsworth, Middlesex, England: Pelican.

David Braun

Department of Philosophy

University of Rochester

Rochester, NY 14627-0078

USA

david.braun@rochester.edu

Jennifer Saul

Department of Philosophy

University of Sheffield

Sheffield S10 2TN

United Kingdom

j.saul@sheffield.ac.uk

Endnotes

¹ The authors contributed equally to the writing of this paper. Their names appear in alphabetical order.

² When we describe the standard intuitions about these sentences, we are speaking of intuitions about whether the sentences are true in the Superman fiction. We think that this complication makes no substantive differences to what follows. In any case, as Saul (1997) points out, there are pairs of sentences entirely about real things that provoke the same intuitions and raise similar issues, for instance (i)-(ii):

- (i) Andropov often visited Leningrad, but never visited St. Petersburg.
- (ii) Andropov often visited Leningrad, but never visited Leningrad.

³ You may be a particularly wary philosopher, who always thinks long and hard before making any judgments about sentences that appear in philosophy journals (lest you be caught in some intellectual trap). If you are, then you may have paused, thought hard, and then judged that (1) and (1*) are both false, and that (2) and (2*) cannot differ in truth-value. The fact that you had these (somewhat unusual) initial intuitions, if you did, is compatible with our solution to the puzzle that we are about to describe. (It's also compatible with all other proposed solutions to the puzzle.)

⁴ We assume that 'Superman' and 'Clark Kent' are rigid designators.

⁵ Some of the sentences in Saul (1997), as Forbes (1997) and Moore (1999) have noted, arguably involve covert reference to psychological attitudes. For the purposes of this paper, we focus on sentences like (1)-(5*), which clearly don't.

⁶ In a recent paper, David Pitt (2001) has presented a semantic solution to the puzzle that differs in significant respects from those of Forbes and Moore (see note 11 for a description of one important difference). Because of limitations of space, we are unable to discuss his view in detail. We hope to do so in future work.

⁷ At least, we don't accept the most common sort of pragmatic explanation. The term 'pragmatic' gets used in many ways, and there are probably meanings of the term on which our explanation counts as pragmatic.

⁸ More precisely (here we go a bit beyond what Forbes and Moore explicitly say): on their views, some utterances of (1), and some acts of inscribing (1), in some contexts, semantically express true propositions; similarly, some utterances of, and acts of inscribing, sentences (2) and (2*) semantically express propositions that really can differ in truth-value. When a person reads printed tokens of (1)-(2*), she may take them as evidence that the author(s) of the tokens engaged in some act of inscribing in some context. In what follows, we often ignore these details (as do Forbes and Moore themselves). Sometimes we'll speak as though Forbes's and Moore's views entail that readers read certain inscriptions, and that these inscriptions expressed certain propositions. (If 'inscription' here means something like token, then remarks of this sort are, strictly speaking, incorrect; and if 'inscription' means act of inscribing, then, strictly speaking, readers rarely read inscriptions.) See Perry 1997 for more on tokens vs. utterances. See note 22 for further complications.

⁹ Also, utterances of these names may refer to different aspects on different occasions.

¹⁰ Which modes of presentation are involved in the propositions expressed by utterances of (1) may also vary.

¹¹ Pitt's (2001) theory says that all utterances of 'Superman' and 'Clark Kent' refer to distinct alter egos of a single person (Kal El). Thus his theory entails that all utterances of (8) are false. His view similarly entails that all utterances of the identity sentence (3) semantically express a false proposition. We think that these consequences of Pitt's view are incorrect (as are other consequences), but limitations of space prevent us from saying more here.

¹² Forbes and Moore do not mention it, but their views allow that a speaker who is unenlightened about the Superman/Clark double life could use 'Superman' and 'Clark' to speak about aspects or personae of the individual Superman/Clark. For instance, though Lois believes that Superman is not Clark, she may still wish to speak of two aspects of Clark Kent, the diffident bumbling klutz and the aggressive muckraking reporter, and she may sometimes use 'Clark' to refer to one or the other of these aspects. (In this case, Lois would be enlightened about the bumbling klutz/muckraking reporter double life, and so enlightened about the double life relevant to her utterances.) We will ignore this possibility from here on.

¹³ The reason that unenlightened speakers can (nevertheless) think that (1) is true and (1*) is false, and that (2) and (2*) can differ in truth-value, is that they don't believe the proposition that (the individual) Superman is (the individual) Clark Kent.

¹⁴ Here are some details. A crucial premise in the argument of the third paragraph of our introduction is the claim that (1*) follows from (1) and (3). Let's take this to be the claim that the following argument is valid:

(1) Superman leaps more tall buildings than Clark Kent.

(3) Superman is identical with Clark Kent.

(1*) Therefore, Superman leaps more tall buildings than Superman.

Forbes and Moore hold that (1) and (1*) are context-sensitive. Thus, to assess the claim that this argument is valid, they must use a notion of validity that applies to arguments containing context-sensitive sentences. The best such notion is David Kaplan's (1989) (some remarks by Forbes (1999, p. 88) suggest that he would agree). On Kaplan's theory, an argument is valid iff for every context in every model, if the premises are true in that context in that model, then so is the conclusion. (Alternatively: if the premises express propositions in that context that are true in that context, then so does the conclusion.) Given this definition of validity, Moore can say that the argument is valid: in every context in which (3) is true, the names 'Superman' and 'Clark Kent' refer in that context to the same thing (either an ordinary individual or an aspect). In such contexts, the occurrences of the names in (1) and (1*) also refer to the same thing, and so if the premises are true in the context, then the conclusion is also true in the context. However, Moore can say that the context in which you read inscriptions of (1) and (1*) is one in which the names referred to distinct aspects. Thus, (3) was false in the context. (In most contexts in which (3) is uttered, it is true. But (3) was not uttered in this context.) Moore could say that the reasoning of the third paragraph of our introduction went wrong when it (implicitly) assumed that (3) is true in every context, including the context in which you read (1) and (1*). Matters are more complicated for Forbes. According to him, sentences (1) and (1*) are ambiguous. One reading of (1) can be represented by a "logical form" that contains the expression 'so-personified' (or something similar), as does our (1F). But there is another reading of (1) that is properly represented by a logical form which does not contain 'so-personified'. (1*) is similarly ambiguous. The logical

forms in which ‘so-personified’ appears are context-sensitive, for the reference of ‘so’ depends on context. Now consider a disambiguation of the above argument in which both (1) and (1*) are correctly represented by logical forms that include the phrase ‘so-personified’. There are contexts in which (3) and the ‘so-personified’ reading of (1) are true, but in which the ‘so-personified’ reading of (1*) is false. Thus this disambiguated argument is invalid. (On another disambiguation, the argument is valid.) Forbes could say that the reasoning of the third paragraph of our introduction went wrong in two ways: it (implicitly) assumed that the sentences of the argument are unambiguous, and it (implicitly) assumed that they are not context-sensitive. (Thanks to Forbes and Moore for discussion of these matters.)

¹⁵ By contrast, Forbes and Moore say that there are contexts (namely, some enlightened contexts) in which (1) and (1*) express distinct propositions that are not necessarily equivalent.

¹⁶ We are assuming that standard neo-Fregean theories entail that ‘Superman’ and ‘Clark Kent’ are rigid designators, for otherwise they would be vulnerable to standard Kripkean modal objections.

¹⁷ Barber (2000) says that the implicated propositions concern the individual Superman/Clark and the attribute Supermanizing or the attribute Clark Kentizing. We assume here that these attributes are roughly the same as those of occupying the Superman persona and occupying the Clark Kent persona, and so Barber’s implicated propositions are very like Forbes’s and Moore’s semantically expressed propositions. (There are in fact differences, but none that are relevant to this discussion.)

¹⁸ For more on this sort of story, see Salmon 1986, 1989, and Braun 1998, 2002. In our view, all neo-Russellians need to make use of ways of taking propositions; see Braun 2002.

¹⁹ On the neo-Russellian theory, an implicated proposition can be entertained and believed in a variety of ways (or under a variety of guises). The hearer will conclude that (1) is true only if she entertains, and believes, the implicated proposition in the right way. See Braun 1998 and 2002.

²⁰ Grice's characterisation arguably also implies that an utterance of (1) implicates a proposition about aspects only if the audience is enlightened, due to condition (2). But this isn't relevant to our concerns here. For more on the consequences of condition (2), see Saul 2001, 2002.

²¹ A closely related problem for implicature accounts is that the alleged implicatures cannot be either generalized (arising as a default due to certain words, phrases, or constructions) or particularized (depending very strongly on context, rather than arising as a default). Your context did not contain any prior discussion of Superman/Clark's double-life; moreover, you had not been given any information about whether the relevant speaker was thinking about aspects. Thus your intuitions occurred in a context in which there was no contextual support for a particularized implicature concerning aspects. Therefore, a theorist who wishes to explain your intuitions by appeal to implicatures must maintain that the relevant implicatures are generalized. This theorist would then have to say that there are some words, phrases, or constructions that give rise to such implicatures, even in the absence of contextual information. It seems that the only candidates for creating such implicatures are the names 'Superman' and 'Clark

Kent' themselves; so the theorist must claim that utterances of sentences containing these names give rise to implicatures concerning aspects, even in the absence of contextual information. But, as we pointed out in the main text, Grice's theory entails that unenlightened speakers' utterances of sentences containing these names do not implicate propositions about aspects. So it seems that the alleged implicatures concerning aspects cannot be generalized. Some implicature-theorists might want to reject this last aspect of Grice's theory, and maintain that utterances by unenlightened speakers do implicate aspect-propositions. But this reply is problematic. The unenlightened speakers that we are considering (e.g., Lois) understand sentences containing these names, and are otherwise perfectly competent with them. And yet they would never entertain the alleged implicatures when they utter or hear the relevant sentences. In fact, they would be unable to calculate these implicatures. So the reply implies that there are generalized conversational implicatures that some competent speakers cannot calculate and never entertain. This is surely incorrect.

²² We think there are some tricky issues about how sentences like these are supposed to be understood when they are presented in the manner that they are in this paper.

Arguably, the authors are merely mentioning, rather than using, the sentences, and you are meant to evaluate them with respect to some conversation in which they are used rather than mentioned. See also note 8.

²³ Moore (1999, pp. 93-4) considers a similar case, but reports that his intuitions are different from those that we predict. We suspect that the judgments that he is reporting are not his initial ("snap") judgments, and may be influenced by his theory.

²⁴ We are here assuming that you had the standard intuitions when you first encountered the sentences. If you didn't, then you probably did not have the intuition that we predicted you would have about the Lois-Myrtle case. Nevertheless, we believe that many competent, rational, reflective, well-informed readers will have the predicted intuition about Lois's utterance. This is enough to generate the problem for the semantic and implicature accounts that we describe in the main text.

²⁵ Two remarks about (MP): (i) (MP) concerns speakers who judge that U has a possible truth-value T. This is shorthand for talk about speakers who judge that certain sentences could be true (or could be false). We have in mind, for example, speakers who judge that a particular utterance of the conjunction of (2) and the negation of (2*) could be true. (ii) (MP) concerns speakers who are relevantly well-informed. It's difficult to say in a completely general way what this amounts to. But for many of the cases we discuss here, a speaker counts as relevantly well-informed only if she believes certain identities (in the right ways). For example, in cases in which the speaker is making judgments about the truth-values, or possible differences in truth-value, among (1)-(2*), the speaker must believe that Superman is Clark Kent (in a 'Superman'-'Clark' way). Similarly for cases in which the speaker is making judgments about belief sentences, such as 'Lois believes that Superman can fly' and 'Lois believes that Clark can fly'.

²⁶ Some qualifications regarding some neo-Russellians: (i) As we noted in section 3.3, some neo-Russellians, including Salmon and Soames, provide non-semantic, non-implicature explanations of speakers' intuitions about some simple sentences. For instance, Salmon and Soames say that a speaker who accepts 'Superman is Superman' but rejects 'Superman is Clark' does so because she believes the proposition that

Superman is Superman under one guise, but believes its negation under another, suitably different guise. (Recall, however, that an appeal to guises is not sufficient to explain speakers' intuitions about (1)-(2*).) (ii) Despite this, many neo-Russellians, including Salmon and Soames, turn to implicatures (or something like them) to explain speakers' intuitions about belief sentences. Soames (1995) is explicit about this. Salmon (1989) points to the usefulness of Gricean distinctions, but shies away from endorsing the mechanism of implicature. Still, he does not suggest another mechanism. (iii) Thus, some neo-Russellians, including Salmon and Soames, implicitly reject the fully general version of (MP) given in the text, but seem to (tacitly) accept a version of (MP) that is restricted to belief sentences. We suspect that even this restricted version of (MP) is false. See section 5.

²⁷ There are, however, certain logically equivalent tasks on which people perform very well.

²⁸ For a good overview of approaches to the Wason selection task, see Evans, Newstead, and Byrne 1993.

²⁹ The Moses illusion was first studied (and named) in Erickson and Mattson 1981. Subsequent studies include Reder and Kusbit 1991, Kamas, et al. 1996, and Hannon and Daneman 2001.

³⁰ In this type of case, a speaker satisfies (MP)'s requirement of being relevantly well-informed only if she believes (in the right way) that Noah took two animals of each type on to the ark and believes (in the right way) that Moses did not.

³¹ A neo-Fregean could say that you believe that Superman leaps tall buildings, but you don't believe that Clark does, although the propositions you believe (including the

proposition that Superman is Clark) logically imply that Clark leaps tall buildings. See below for more details, and a neo-Russellian version.

³² A third reason: the speculative explanations illustrate the neo-Russellian idea that a person can believe and entertain a single proposition under different guises.

³³ There are good reasons for thinking that English is not your language of thought, such as Kripke's Paderewski case (Kripke 1979). But, for convenience, we ignore these matters here.

³⁴ We believe that Schiffer (1981) first introduced the belief box terminology.

³⁵ Strictly speaking, a person entertains a proposition by having a sentence in his entertainment box. But in what follows, we will often say that a person entertains a sentence, such as (1) or (2). By this we will mean that the person has that sentence in his entertainment box, and thereby entertains the proposition expressed by the sentence.

³⁶ Our conception of a file is similar to Perry's (1980). Perry, however, tends to think of files as (something like) collections of open formulas, whereas we tend to think of them as collections of closed sentences containing names.

³⁷ For simplicity, we will often write of 'double lives', but not all simple sentence puzzle cases involve double lives—for example, some such cases involve 'St Petersburg' and 'Leningrad'. Our explanation is meant to apply to these other cases as well.

³⁸ Presumably, the sentence 'Superman is Clark' is present in each person's 'Superman' file and each person's 'Clark' file. But see note 39 below.

³⁹ This is so even though the identity sentence 'Superman is Clark' is (presumably) present in a typical enlightened person's 'Superman' file and 'Clark' file. One reason a typical enlightened speaker may (nevertheless) fail to make the substitution inference is

that the identity sentence may remain inactive when she opens her ‘Superman’ and ‘Clark’ files, or may become less active than other sentences in the files. (This low level of activity may, in turn, be due to the fact that typical speakers think less often about Superman/Clark’s double life than they do about his other attributes.) If so, then the identity sentence will not be entertained, and the substitution inferences that it licenses will not be performed.

⁴⁰ There are vexing questions about accuracy conditions for images that we won’t try to decide here. Consider the image you formed when you heard (2). Is it accurate if some bespectacled man (of a certain appearance) went into a phone booth and some caped man (of a certain appearance) came out? Or does its accuracy require that Superman/Clark himself, while bespectacled and appearing a certain way, went into a phone booth, and that Superman/Clark himself, while becaped and appearing a certain way, came out? In short, are the accuracy conditions of images general or singular? We don’t know. But it doesn’t matter for our purposes, because, whether singular or not, these images do differ in their accuracy conditions.

⁴¹ This may be the sort of explanation that Ostertag (1998) has in mind.

⁴² Of course, the ordinary speaker's retraction is also compatible with the semantic and implicature theories. These theories could claim that the coaching results in a change in context that gives rise to different propositions either semantically expressed or conversationally implicated. But we have already given reasons to reject those views.

⁴³ Of course, identity sentences like (3) are also standard philosophical examples that contain occurrences of two names, but they don’t provoke strong anti-substitution intuitions. We say below (and in note 44) why we think they do not.

⁴⁴ If the sentence that you are considering is an identity sentence containing two names, then the unconscious presumption of non-coreference either never arises or is immediately overridden.

⁴⁵ For Levinson's work on pragmatics and anaphora, see Levinson 2000, especially chapter 4. For an introduction to Discourse Representation Theory, see Kamp and Reyle 1993. An explanation along Levinson's lines would claim that an utterance of (1) triggers a generalized conversational implicature that the names 'Superman' and 'Clark Kent' fail to co-refer. Levinson thinks that such implicatures are default, overridable assumptions or inferences. Such Levinsonian implicatures differ from Barber's implicatures in two respects. First, the Levinsonian implicatures concern co-reference, whereas Barber's implicatures concern aspects. Second, the Levinsonian implicatures are false, and thus do not satisfy (MP), whereas Barber's implicatures are supposed to be true and satisfy (MP). One might wonder whether syntactic constraints on referential dependence (e.g., Chomsky's Binding Condition C on co-indexation) play some role in creating the presumption of non-coreference. We think not, for the two names in sentence (2) occur in two different clauses that are joined by conjunction; it's thus unlikely that a syntactic theory would entail anything about their referential dependence, or lack thereof. (Thanks to Jeffrey Runner for discussion of this last point.)

⁴⁶ Braun's theory does not say that speakers' belief-sentence intuitions are due to the sorts of mistakes that we think occur during typical speakers' evaluations of (1)-(2*). In fact, one of us (DB) thinks that ordinary speakers resist substitution more strongly in belief sentences than in simple sentences, and that this is evidence that our explanation of simple-sentence intuitions cannot be extended to belief sentences.

⁴⁷ Some readers of Saul (1997) might think that she endorsed the following argument.

She did not. See section 2 for a description of her argument in that paper.

⁴⁸ As far as we know, Braun's (1998) theory and Soames's (2002) theory are the only ones. Braun uses neither semantics nor implicatures to explain anti-substitution intuitions about belief sentences. On his theory, speakers who have these intuitions usually do not entertain propositions whose truth-values match their intuitions. According to Soames, speakers who utter belief sentences containing names (such as 'Lois believes that Superman flies' and 'Lois does not believe that Clark flies') usually *assert* propositions that ascribe belief (or lack of belief) in descriptive propositions (e.g., the proposition that Lois believes that the person who wears a red cape, Superman, flies, and the proposition that Lois does not believe that the bespectacled reporter, Clark, flies). Soames says that these asserted propositions are neither semantically expressed nor conversationally implicated, but do explain anti-substitution intuitions. Thus, Soames's theory is also inconsistent with (MP). However, it is much closer in spirit to (MP) than Braun's theory. In particular, it assumes that intuitions are to be explained by propositions whose truth-values match those intuitions—only these propositions are asserted rather than semantically expressed or conversationally implicated. For some critical discussion, see Braun 2002.

⁴⁹ Thanks to Paul Pietroski for helping us to see this.

⁵⁰ Thanks to Alex Barber, John G. Bennett, Ben Caplan, Graeme Forbes, Christopher Hookway, David Hunter, Gail Mauner, Joseph Moore, Gary Ostertag, Paul Pietroski, Stefano Predelli, Teresa Robertson, Jeffrey Runner, the participants at Eros Corrazza's *On Referring* conference, the participants at the September 2001 session of Ernest

Lepore's semantics workshop, and an anonymous referee for *Philosophical Studies*, for many useful comments and discussions. Thanks especially to Paul Pietroski, who was our commentator at the Lepore workshop. David Braun wishes to give special thanks to Gail Mauner for making him aware of the Moses illusion. Jenny Saul would like to thank Chris Hookway in particular. Discussions with and suggestions from him motivated her to pursue a solution like the one proposed here, and have strongly influenced her thinking about these issues.