2

Reflexivity (1986)

In 1983 Mark Richard formulated a new and interesting problem for theories of direct reference with regard to propositional–attitude attributions.[1] The problem was later discovered independently by Scott Soames, who recently advanced it[2] as a powerful objection to the theory put forward by Jon Barwise and Jon Perry in *Situations and Attitudes*.[3] Interestingly, although both Richard and Soames advocate the fundamental assumption on which their philosophical problem arises, they disagree concerning the correct solution to the problem. In this paper I discuss the Richard–Soames problem, as I shall call it, as well as certain related problems and puzzles involving reflexive constructions in propositional–attitude attributions. I will treat these problems by applying ideas I invoked in *Frege's Puzzle*[4] defending a semantic theory that shares certain features with, but differs significantly from, that of Barwise and Perry. Unlike the theory of *Situations and Attitudes*, the theory of *Frege's Puzzle* has the resources without modification to solve the Richard–Soames problem and related problems.

I

In setting out the Richard–Soames problem, we make some important assumptions. First, we make the relatively uncontroversial assumption that a monadic predicate ⌜believes that S⌝, where S is a declarative sentence, is simply the result of filling the second argument place of the dyadic, fully extensional predicate 'believes' with the term ⌜that S⌝. Furthermore, it is assumed that the contribution made by the dyadic predicate 'believes' to securing the information content (with respect to a time t) of, or the proposition expressed (with respect to t) by, a declarative sentence in which the

Many of the ideas in this paper were first urged by me in correspondence with David Kaplan, Mark Richard, and Scott Soames in February 1984. There was also a discussion of some of these issues with Joseph Almog, Kaplan, and Soames, and some later correspondence with Alonzo Church. Although there was not the time before submission to receive reactions or comments on the present paper, it has benefited from these earlier exchanges.

[1] M. Richard, 'Direct Reference and Ascriptions of Belief', *Journal of Philosophical Logic* 12 (1983), 425–452; also in this volume.

[2] 'Lost innocence', *Linguistics and Philosophy* 8 (1985), 59–71.

[3] MIT Press, 1983.

[4] MIT Press, 1986.

predicate occurs (outside of the scope of any nonextensional devices, such as quotation marks) is a certain binary relation between believers and propositions, the relation of believing–at–t,[5] and that a term of the form ⌜that S⌝ refers (with respect to a possible context of use c) to the information content (with respect to c) of the sentence S itself. More accurately, the following is assumed:

(*B*) A monadic predicate of the form ⌜believes that S⌝, where S is an (open or closed) sentence, correctly applies (with respect to a possible context of use and an assignment of values to individual variables) to all and only those individuals who stand in the binary belief relation (at the time of the context in the possible world of the context) to the information content of, or the proposition expressed by, S (with respect to that context and assignment).

On this assumption, a sentence of the form ⌜a believes that S⌝, where a is any singular term, is true if and only if the referent of a stands in the belief relation to the information content of S. Thesis (*B*) is generally agreed upon by Fregeans and Russellians alike, and is more or less a commonplace in the literature of the theory of meaning, and of the philosophy of semantics generally.

In addition to thesis (*B*), we assume that ordinary proper names, demonstratives, other single-word indexicals (such as 'he'), and other simple (noncompound) singular terms are, in a given possible context of use, Russellian "genuine names in the strict logical sense".[6] Put more fully, we assume the following anti-Fregean thesis as a hypothesis:

(*R*) The contribution made by an ordinary proper name, demonstrative, or other simple singular term to securing the information content of, or the proposition expressed by, declarative sentences (with respect to a given possible context of use) in which the term occurs (outside of the scope of nonextensional operators, such as quotation marks) is just the referent of the term, or the bearer of the name (with respect to that context of use).

In various alternative terminologies, it is assumed that the *interpretation* (Barwise and Perry), or the *Erkenntniswerte* (Frege), or the *content* (David Kaplan), or the *meaning* (Russell), or the *semantic value* (Soames), or the *information value* (myself) of a proper name, demonstrative, or other simple singular term, with respect to a given context, is just its referent.

It is well-known that the thesis that ordinary proper names are Russellian, in this sense, in conjunction with thesis (*B*), gives rise to problems in propositional–attitude attributions, and is consequently relatively unpopular. (Even Russell rejected it.) Thus, thesis (*R*) is hardly the sort of thesis that can legitimately be taken for granted as accepted by the reader. However, I defend thesis (*R*) at some length and in some detail

[5] The idea of indexing, or relativizing, the notion of information content to times (independently of contexts) is due to M. Richard, 'Tense, Propositions, and Meanings', *Philosophical Studies* 41 (1982), 337–351. The idea that the contribution made by a predicate to information content is something like a temporally indexed attribute is defended in *Frege's Puzzle* and stems from Richard's idea of indexing information content to times.

[6] B. Russell, 'The Philosophy of Logical Atomism', in R. C. Marsh (ed.), *Logic and Knowledge* (George Allen and Unwin: London, 1956), 177–281; also pp. 35–155 in Russell's *The Philosophy of Logical Atomism*, ed. D. Pears (Open Court: La Salle, 1985).

in *Frege's Puzzle*. Moreover, the thesis has gained some long overdue respectability recently, and it cannot be summarily dismissed as obviously misguided. It is (more or less) accepted by Barwise–Perry, Kaplan, Richard, Soames, and others. One standard argument against the thesis—the argument from apparent failure of substitutivity in propositional–attitude contexts—has been shown by Kripke[7] to be inconclusive at best, and the major rival approaches to the semantics of proper names and other simple singular terms have been essentially refuted by Keith Donnellan, Kripke, Perry, and others.[8] The Richard–Soames problem is a problem that arises only on the assumption of thesis (*R*), and it is a problem for this thesis. It is not a problem for alternative approaches, such as those of Frege or Russell, which have much more serious problems of their own. Thesis (*R*) is to be taken as a hypothesis of the present paper, its defence given elsewhere. The conclusions and results reached in the present paper on the assumption of thesis (*R*) may be regarded as having the form 'If thesis (*R*) is true, then thus-and-so.' The present paper, in combination with *Frege's Puzzle*, allows for the all-important *modus ponens* step.

FN:7

FN:8

One version of the Richard–Soames problem can be demonstrated by the following sort of example, derived from Richard's. Suppose that Lois Lane, who is on holiday somewhere in the wilderness, happens to overhear an elaborate plot by some villainous misanthrope to expose Superman to Kryptonite (the only known substance that can harm Superman) at the Metropolis Centennial Parade tomorrow. She quickly rushes for the nearest telephone to warn Superman, but suddenly remembers that the nearest telephone is one day's journey away. As luck would have it, she happens to be standing in front of an overnight mail delivery service outlet. She quickly scribbles a note warning of the plot to harm Superman—a note that absolutely, positively has to get there overnight. She has no address for Superman (or so she believes), but she does have Clark Kent's address, and she (thinks she) knows that Clark planned to spend all day tomorrow at his flat. Now the following sentence is true:

(1*a*) Lois believes that she will directly inform Clark Kent of Superman's danger with her note.

By the assumption of theses (*B*) and (*R*), it would seem that the following sentence contains the very same information as (1*a*), and hence must be true as well:

(1*b*) Lois believes that she will directly inform Superman of Superman's danger with her note.

[7] S. Kripke, 'A Puzzle about Belief' in A. Margalit (ed.) *Meaning and Use*, (D. Reidel: Dordrecht, 1979), 239–275; also in this volume.

[8] See K. Donnellan 'Proper Names and Identifying Descriptions', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (D. Reidel: Dordrecht, 1972), 356–379; S. Kripke *Naming and Necessity* (Harvard University Press and Basil Blackwell, 1972, 1980); also in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, (D. Reidel: Dordrecht, 1972), 253–355, 763–769; and J. Perry 'The Problem of the Essential Indexical', *Nous* 13 (1979), 3–21; also in this volume. For a summary of the major difficulties with the views of Frege and Russell, see N. Salmon, *Reference and Essence* (Princeton University Press and Basil Blackwell, 1981), chapter 1. Further problems with the Frege–Russell view in connection with propositional attitudes are discussed in *Frege's Puzzle*, ch. 9 and *passim*.

Richard argues, however, that although (1*a*) is true in this example, (1*b*) cannot be true. For if (1*b*) were true, then the following sentence would also be true:

(1*c*) Lois believes that there is someone *x* such that she will directly inform *x* of *x*'s danger with her note.

That is, if (1*b*) were true, then Lois would also believe that someone or other is such that she will inform him of *his own* danger with her note, since this follows trivially by existential generalization from what she believes according to (1*b*). Yet Lois believes no such thing. (Recall that Lois believes that she has no address for Superman.) Of course, Lois hopes that Clark will relay the warning to Superman before it is too late, but she has not formed the opinion that she herself will directly inform someone of his own danger with her note. To put it another way, it is simply false that Lois believes that there is someone with the special property that he will be directly informed by her of his own danger with her note. On the contrary, what she believes is that she will inform someone of *someone else's* danger with her note. Thus (1*a*) is true, though (1*b*) would seem to be false. This poses a serious problem for any theory—such as the theory formed from thesis (*R*) coupled with thesis (*B*) and some other natural assumptions—that claims that (1*a*) and (1*b*) have exactly the same information content, or even merely that they have the same truth value.

Using a similar example, Soames provides a powerful argument against semantic theories of a type that identify the information contents of declarative sentences with sets of *circumstances* (of some sort or other) with respect to which those sentences are either true or untrue (or equivalently, with characteristic functions from circumstances to truth values)—such as the possible-world theories of information content (David Lewis, Robert Stalnaker, and many others) or the "situation" theory of *Situations and Attitudes*. The argument is this: the following sentence concerning a particular ancient astronomer is assumed to be true (where reference to a language, such as 'English', is suppressed):

(2*a*) The astronomer believes: that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus.

Hence according to thesis (*R*) in conjunction with thesis (*B*) and some natural assumptions, the following sentence, which allegedly contains the very same information as (2*a*), must also be true:

(2*b*) The astronomer believes: that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Hesperus.

But if (2*b*) is true, and thesis (*B*) is also true, then on certain assumptions that are either trivial or fundamental to a set-of-circumstances theory of information content, the following is also true:

(2*c*) The astronomer believes: that something or other is such that 'Hesperus' refers to it and 'Phosphorus' refers to it.

Assuming thesis (*B*), the additional assumptions needed to validate the move from (2*b*) to (2*c*) on any set-of-circumstances theory of information content are: (i) that

a believer's beliefs are closed under *simplification* inferences from a conjunction to either of its conjuncts, i.e. if *x* believes *p and q*, then *x* believes *q*; and (ii) that the conjunction of an ordinary sentence *S* (excluding nonreferring singular terms and nonextensional devices such as the predicate 'does not exist') and any existential generalization of *S* is true with respect to exactly the same circumstances as *S* itself.

Now (2*c*) is tantamount to the claim that the astronomer believes that 'Hesperus' and 'Phosphorus' are co-referential. Yet certainly (2*c*) is no consequence of (2*a*). Indeed, we may take it as an additional hypothesis that (2*c*) is false of the ancient astronomer in question. Since (2*a*) is true and (2*c*) is false, it is either false that if (2*a*) then (2*b*)—contrary to the conjunction of theses (*B*) and (*R*)—or else it is false that if (2*b*) then (2*c*)—contrary to the conjunction of (*B*) and any set-of-circumstances theory of information content. Now (*B*) and (*R*) are true Therefore, Soames argues, any set-of-circumstances theory of information content is incorrect. As Soames points out, the problem points to a fundamental error in the theory of *Situations and Attitudes*, which accepts both (*B*) and (*R*) as fundamental, thereby ensuring the validity of the move from (2*a*) to (2*b*), as well as the assumptions that validate the move from (2*b*) to (2*c*).

In the general case, we may have the first of the following three sentences true and the third false, where *a* and *b* are co-referential proper names, demonstratives, other simple singular terms, or any combination thereof, and *R* is a dyadic predicate:

(3*a*) *c* believes that *aRb*

(3*b*) *c* believes that *aRa*

(3*c*) *c* believes that $(\exists x)xRx$.

The Richard–Soames problem is that (3*b*) appears to follow from (3*a*), and (3*c*) appears to follow from (3*b*). Since (3*a*) is true and (3*c*) false, something has got to give.

II

Now (3*b*) is either true or false. Hence it is either false that if (3*a*) then (3*b*), or else it is false that if (3*b*) then (3*c*). Both Richard and Soames accept thesis (*R*). Insisting that if (3*b*) then (3*c*), Richard maintains that it is false that if (3*a*) then (3*b*), thereby impugning thesis (*B*).[9] Accepting thesis (*B*) as well as (*R*), Soames argues instead that "there is a principled means of blocking" the move from (3*b*) to (3*c*) while preserving (*B*).

FN:9

There is a certain intuitive picture of belief advanced by Barwise and Perry (Chapter 10) and which is independently plausible in its own right. This is a picture of belief as a cognitive state arising from internal mental states that derive information content in part from causal relations to external objects. Soames points out that on this picture of belief, the following is indeed true if (3*b*) is:

(3*d*) $(\exists x)$ *c* believes that *xRx*.

[9] 'Direct Reference and Ascriptions of Belief'; pp. 440–2 and *passim*. He constructs (pp. 444–445) a semantics for belief attributions that conflicts with thesis (*B*).

FN:10 Soames adds:[10]

> However, [on this picture of belief] there is no reason to think that [the referent of *c*] believes the proposition that something bears *R* to itself. Since none of the agent's mental states has this as its information content, he does not believe it.

Quine distinguishes two readings of any sentence of the form ⌜*c* believes something is ϕ⌝—what he calls the *notional* and the *relational* readings. The notional reading may be spelt out as ⌜*c* believes: that something or other is ϕ⌝. It is the Russellian secondary occurrence or small–scope reading. The relational reading may be spelt out as ⌜*c* believes something in particular to be ϕ⌝, or more perspicuously as ⌜Something is such that *c* believes: that it is ϕ⌝. It is the Russellian primary occurrence or large–scope reading. In Quine's terminology, Soames claims that the notional reading of ⌜*c* believes something bears *R* to it⌝ does not follow from the relational. Quine demonstrated some time ago that the relational reading of ⌜*c* believes something is ϕ⌝ does not in general follow from the notional reading, with his clever example of 'Ralph believes someone is a spy'. Soames may be seen as arguing that, on a certain plausible picture of belief, there are cases in which the reverse inference also fails. Since the appearance of Quine's influential writings on the subject, it is no longer surprising that the notional reading does not imply the relational. It is at least somewhat surprising, however, that there could be converse cases in which the relational reading is true yet the notional reading false. This is what Soames is arguing.

My own view of the Richard–Soames problem favours Soames's account over Richard's. Thesis (*B*) is supported by strong linguistic evidence. It provides the simplest and most plausible explanation, for example, of the validity of such inferences as:

> John believes the proposition to which our nation is dedicated.
> Our nation is dedicated to the proposition that all men are created equal.
> Therefore, John believes that all men are created equal.

Furthermore, although a number of philosophers have proposed a variety of truth-condition assignments for belief attributions contrary to thesis (*B*), these alternative truth-condition assignments often falter with respect to belief attributions that involve open sentences as their complement 'that'-clause, and that are true under some particular assignment of values to individual variables or to pronouns—for example, 'the astronomer believes that *x* is a planet' in 'There is something *x* such that *x* = Venus and the astronomer believes that *x* is a planet' or 'the astronomer believes that it is a planet' in 'As regards Venus, the astronomer believes that it is a planet'.[11] Thesis (*B*) should be maintained to the extent that the facts allow, and should not be abandoned if Soames is correct that there is a principled means of solving the Richard–Soames problem while maintaining (*B*).

FN:11

By contrast, Soames's proposals for solving the problem invoke essentially some of the same ideas advanced and defended in *Frege's Puzzle*. There I develop and defend

[10] 'Lost Innocence', p. 62.

[11] Richard 'Direct Reference and Ascriptions of Belief' is one exception.

thesis (*R*) (and, to a lesser extent, thesis (*B*)), as well as the view (which Russell himself came to reject) that the contents of beliefs formulatable using ordinary proper names, demonstratives, or other simple singular terms, are so-called singular propositions (Kaplan), i.e. structured propositions directly about some individual which occurs directly as a constituent of the proposition. I take propositions to be structured in such a way that the structure and constituents of a proposition are directly readable from the structure and constituents of a declarative sentence containing the proposition as its information content. By and large, a simple (noncompound) expression contributes a single entity, taken as a simple (noncomplex) unit, to the information content of a sentence in which the expression occurs, whereas the contribution of a compound expression (such as a phrase or sentential component) is a complex entity composed of the contributions of the simple components.[12] One consequence of this sort of theory is that, contrary to set-of-circumstances theories of information content, there is a difference, and therefore a distinction, between the information content of the conjunction of an ordinary sentence *S* and any of its existential generalizations and that of *S* itself. This disables the argument that applied in the case of a set-of-circumstances theory to establish the (alleged) validity of the move from (3*b*) to (3*c*).

FN:12

Unfortunately, this difference between the two sorts of theories of information content does not make the problem disappear altogether. There is an interesting philosophical puzzle concerning the logic and semantics of propositional–attitude attributions that is generated by the Richard–Soames problem, a puzzle that arises even on the structured-singular-proposition sort of view sketched above.

Soames slightly misstates the case when he says that (on the intuitive picture of belief as deriving from certain mental states having information content), 'there is no reason to think that (3*c*) is true'. For in fact, even though (1*c*) and (2*c*) are false in the above examples, there are very good reasons to think that they are true. One excellent reason to think that (1*c*) is true is the fact that (1*b*) is true, and one excellent reason to think that (2*c*) is true is the fact that (2*b*) is true. In general, it is to be expected that if a sentence of the form $\ulcorner c$ believes that $\phi_a \urcorner$ is true, then so is $\ulcorner c$ believes that $(\exists x)\phi_x \urcorner$, where a is a singular term that refers to something, ϕ is an ordinary extensional context (excluding predicates such as 'does not exist'), and ϕ_a is the result of substituting (free) occurrences of a for free occurrences of 'x' uniformly throughout ϕ_x. There is a general psychological law to the effect that subjects typically tend to believe the existential generalizations of their beliefs. Herein the puzzle arises. Even if the conjunctive proposition '*Hesperus' and 'Phosphorus' refer to Hesperus and*

[12] The reason for the phrase 'by and large' is that there are important classes of exceptions to the general rule. Certain nonextensional operators, such as quotation marks, create contexts in which compound expressions contribute themselves as units to the information content of sentences in which the quotation occurs, and other nonextensional operators, such as temporal operators, create contexts in which some compound expressions contribute complexes other than their customary contribution to information content (see n. 5, above). In addition, we shall see below that a compound predicate formed by abstraction from an open sentence is regarded as contributing something like an attribute, taken as unit, rather than a complex made up of the typical contributions of the compound's components.

there is something that 'Hesperus' and 'Phosphorus' refer to is not the same proposition as the simpler proposition '*Hesperus' and 'Phosphorus' refer to Hesperus*, if the astronomer believes that 'Hesperus' and 'Phosphorus' refer to Hesperus, then it seems he ought to believe that there is something that 'Hesperus' and 'Phosphorus' refer to. And if Lois believes that she will inform Superman of his danger with her note, then it seems she ought to believe that there is someone whom she will inform of his danger with her note. It is precisely for this reason that Richard rejects (1*b*), even though he does not endorse a set-of-circumstances theory of information content and favours the structured-singular-proposition account.

Perhaps if a subject is insane or otherwise severely mentally defective, he or she may fail to believe the (validly derivable) existential generalizations of his or her beliefs, but we may suppose that neither Lois Lane nor the astronomer suffer from any mental defects. We may even suppose that they are master logicians, or worse yet, that they have a perverse penchant for drawing existential generalization (EG) inferences as often as possible. They go around saying things like 'I'm tired now; hence, sometimes someone or other is tired' and 'Fred shaves Fred; hence someone shaves Fred, Fred shaves someone, and someone shaves himself.' In this way, it can be built into the example that the truth of (1*b*) is an excellent reason to believe in the truth of (1*c*), and the truth of (2*b*) is an excellent reason to believe in the truth of (2*c*). For such EG-maniacs, one might expect that it is something of a general law that every instance of the following schema is true:

(L_1) If *c* believes that ϕ_a, then *c* believes that something such that ϕ_{it},

where *c* refers to the subject, *a* is any referring singular term of English, ϕ_{it} is any English sentence in which the pronoun 'it' occurs (free and not in the scope of quotation marks, an existence predicate, or other such operators) and which may also contain occurrences of *a*, and ϕ_a is the result of substituting (free) occurrences of *a* for (free) occurrences of 'it' throughout ϕ_{it}. In fact, one might expect that it is something of a general law that every instance of (L_1) is true where *c* refers to any normal speaker of English, even if he or she is not an EG-maniac.

I maintain with Soames that the sentences ⌜If (1*b*) then (1*c*)⌝ and ⌜If (2*b*) then (2*c*)⌝ constitute genuine counterexamples to this alleged general law. But even if the principle that every instance of (L_1), as formulated, is true is thereby refuted, surely something very much like it, some weakened version of it, *must be* true—even where the referent of *c* does not have a perverse penchant for existential generalization. For the most part, in the typical kind of case, it would be highly irrational for someone to fail to believe the existential generalizations of one of his or her beliefs. Neither Lois Lane nor the astronomer is irrational in this way. The conditionals ⌜If (1*b*) then (1*c*)⌝ and ⌜If (2*b*) then (2*c*)⌝ are not typical instances of schema (L_1), but it is not enough simply to point out how they are atypical and to leave the matter at that. It is incumbent on the philosopher who claims that these instances of (L_1) fail, to offer some alternative principle that is *not* falsified in these cases and thereby accounts for the defeasible reliability, and the *prima facie* plausibility, of the alleged general law.

This is not a problem special to set-of-circumstances theories of information content. It is equally a puzzle for the structured-singular-proposition sort of theory that

I advocate and that Soames proposes in his discussion of the Richard–Soames problem. It is a puzzle for the conjunction of theses (*B*) and (*R*), irrespective of how these theses are supplemented with a theory of information content.

III

There is a second, and surprisingly strong, reason to suppose that (1*c*) and (2*c*) are true. The general puzzle posed by the Richard–Soames problem can be significantly strengthened if we exploit a simple reflexive device already present to a certain degree in standard English.

Given any simple dyadic predicate Π, we may form a monadic pedicate ⌜ *self*-Π ⌝ defined by

$$(\lambda x)x\Pi x,$$

in such a way that ⌜ *self*-Π ⌝ is to be regarded as a simple (noncompound) expression, a single word. In English, this might be accomplished by converting a present tensed transitive verb *V* into a corresponding adjective and prefixing 'self-' to obtain a reflexive adjective; e.g. from 'cleans' we obtain 'self-cleaning', from 'indulges', 'self-indulgent', from 'explains', 'self-explanatory', and so on. The contribution made by a term of the form ⌜ *self*-Π ⌝ to the information content, with respect to a time *t*, of a typical sentence in which it occurs is simply the reflexive property of bearing *R* to oneself at *t*, where *R* is the binary relation semantically associated with Π.[13] Assuming thesis (*R*), if *a* is a proper name or other simple singular term and *R* is the binary relation semantically associated with Π, then the information content, with respect to *t*, of the sentence ⌜ *self*-Π(*a*) ⌝ is the singular proposition made up of the referent of *a* together with the property of bearing *R* to oneself at *t*.

FN:13

Consider again the move from (3*a*) to (3*b*), where *a* and *b* are co-referential proper names, *R* is a simple dyadic predicate, and (3*a*) is true:

(3*a*) *c* believes that *aRb*.

(3*b*) *c* believes that *aRa*.

As Soames points out, (on a plausible picture of belief) the following relational, or *de re*, attribution follows from (3*b*):

(3*d*) (∃*x*) *c* believes that *xRx*.

[13] The '*self*'-prefix defined here may not correspond exactly to that of ordinary English. In English, the term 'self-cleaning' may apply, with respect to a time *t*, to an object even if that object is not cleaning itself *at t* (say, because it is unplugged or switched off for the moment), as long as the object is the *sort* of thing at *t* that cleans itself at appropriate times. Similarly, someone is self-indulgent at *t* if and only if he or she is the sort of person at *t* that has at some appropriate times the feature of indulging oneself, even if he or she is not doing so at *t*. The '*self*'-prefix defined here is such that ⌜ *self*-*R* ⌝ applies to an object with respect to *t* if and only if the object bears *R* to itself at *t*.

In fact, a somewhat stronger *de re* attribution also follows from (3*b*), by exportation:[14] FN:14

($\exists x$) [$x = a$ & c believes that xRx],

or less formally:

(3*b*′) c believes of a that it R it.

Now from this it would seem to follow that:

(3*e*′) c believes of a that it R itself.

From this (perhaps together with some general psychological law) it would seem to follow further that:

(3*f*′) c believes of a that *self-R*(it),

with the predicate ⌜ *self-R* ⌝ understood as explained above. Finally by importation, we may infer:

(3*f*) c believes that *self-R(a)*.

For example, suppose that, owing to certain miscalculations, the astronomer comes to believe that Hesperus weighs at least one thousand tons more than Phosphorus. Now every step in the following derivation follows by an inference pattern that is either at least apparently intuitively valid or else sanctioned by the conjunction of theses (*B*) and (*R*), or both:

(4*a*) The astronomer believes that Hesperus outweighs Phosphorus.

(4*b*) The astronomer believes that Hesperus outweighs Hesperus.

(4*b*′) The astronomer believes of Hesperus that it outweighs it.

(4*e*′) The astronomer believes of Hesperus that it outweighs itself.

(4*f*′) The astronomer believes of Hesperus that it is self-outweighing.

(4*f*) The astronomer believes that Hesperus is self-outweighing.

One could continue the sequence of inferences from (4*f*) all the way to:

(4*c*) The astronomer believes that there is something such that it outweighs it.

[14] The unrestricted rule of exportation has been shown invalid, or at least highly suspect, by the fallacious inference 'The shortest spy exists and Ralph believes: that the shortest spy is a spy; therefore Ralph believes of the shortest spy: that he or she is a spy'. From the conclusion of this inference one may validly infer 'There is someone whom Ralph believes to be a spy', which intuitively does not follow from the initial premiss. This instance of exportation fails because the exported term, 'the shortest spy', is a definite description. The theory formed from the conjunction of theses (*B*) and (*R*) requires the validity of exportation with respect to belief attributions, provided the rule is restricted to proper names, demonstratives, or other simple singular terms. Hence, the theory must accept the inference from (3*b*) to (3*b*′) (assuming the tacit premiss ⌜($\exists x$)[$x = a$]⌝). Similarly, the theory is committed to the validity of importation, inferring ⌜c believes: that ϕ_a⌝ from ⌜c believes of a: that ϕ_{it}⌝, under the same restriction on a.

by invoking some corrected, weakened version of the law mentioned above (the alleged law that every appropriate instance of (L_1) is true), to pass from (4*f*) to:

(4*g*) The astronomer believes that there is something such that it is self-outweighing,

from which (4*c*) appears to follow directly. But there is no need to extend the derivation this far. A problem arises at least as soon as (4*f*). For unless the astronomer is insane, or otherwise severely mentally defective, (4*f*) is obviously false. The astronomer would not ascribe to Venus the reflexive property, which nothing could possibly have, of weighing more than oneself. Hence, in moving from a sentence to its immediate successor, somewhere in the derivation of (4*f*) we move from a truth to a falsehood. Where? The moves from (4*a*) to (4*b*′) and from (4*f*′) to (4*f*) are validated by the conjunction of theses (*B*) and (*R*), and both of the remaining transitions commencing with (4*b*′) are based on inference patterns that (assuming ordinary folk psychology and that the astronomer is normal) seem intuitively valid.

One may harbour some residual doubts about the exportation move from (4*b*) to (4*b*′) and/or the importation move from (4*f*′) to (4*f*). The theory formed from the conjunction of theses (*B*) and (*R*) requires the validity of both of these inferences, so that if either is invalid the theory is false. In fact, however, these inferences are not essential to the present puzzle. The exportation inference takes us on a detour that some may find helpful, though one may bypass the *de re* 'believes of' construction altogether. Instead, we may construct the following alternative derivation from (4*b*):

(4*b*) The astronomer believes that Hesperus outweighs Hesperus.

(4*e*) The astronomer believes that Hesperus outweighs itself.

(4*f*) The astronomer believes that Hesperus is self-outweighing.

If the inference from (4*b*′) to (4*e*′) is valid, then by parity of reasoning so is the inference from (4*b*) to (4*e*). And if the inference from (4*e*′) to (4*f*′) is valid, then by parity of reasoning so is the inference from (4*e*) to (4*f*). Hence, if the derivation of (4*f*′) from (4*b*′) via (4*e*′) is legitimate, then so is the derivation of (4*f*) from (4*b*) via (4*e*). But (4*f*) is false. Therefore, it would seem, so is (4*b*). Sentence (4*a*), on the other hand, is true. This raises anew doubts about the independently suspicious move from (4*a*) to (4*b*), or more generally, the move from (3*a*) to (3*b*), thereby impugning once again the conjunction of theses (*R*) and (*B*).

The new puzzle, then, is this: according to the conjunction of theses (*B*) and (*R*), (4*b*) follows from (4*a*) together with the fact that 'Hesperus' and 'Phosphorus' are co-referential proper names Now in the sequence ⟨(4*b*), (4*e*), (4*f*)⟩, each sentence appears to follow logically from its immediate predecessor. Alternatively in the sequence ⟨(4*b*′), (4*e*′), (4*f*′)⟩, each sentence appears to follow logically from its immediate predecessor, and furthermore, according to the conjunction of (*B*) and (*R*), (4*b*) entails (4*b*′), and (4*f*′) entails (4*f*). One way or another, we seem to be able to derive (4*f*) from (4*a*), together with the fact that 'Hesperus' and 'Phosphorus' are co-referential proper names. Yet in the example, (4*a*) is plainly true and (4*f*) plainly false. Where does the derivation go wrong?

I call this *the puzzle of reflexives in propositional attitudes*. Here again, the problem posed by the puzzle is especially pressing for any set-of-circumstances theory of information content. In fact, the problem is even more pressing than the Richard–Soames problem for such theories, if that is possible. One difference between the Richard–Soames problem and the puzzle of reflexives in propositional attitudes is that what is said to be believed at the final step of the derivation, in this case step (3*f*), is not merely a consequence of, but is *equivalent to*, what is said to be believed in (3*b*). In fact, any circumstance in which an individual x bears R to x is a circumstance in which x has the reflexive property of bearing R to oneself, and vice versa. There is no need here to make the additional assumption that belief is closed under simplification inferences. Any set-of-circumstances theory of information content, in conjunction with thesis (*B*), automatically validates the derivation of (3*f*) from (3*b*). The problem thus also points to a fundamental error in the theory of *Situations and Attitudes* which includes both theses (*B*) and (*R*) as fundamental, thereby validating the full derivation of (3*f*) from (3*a*) without any further assumptions concerning belief. The puzzle of reflexives in propositional attitudes, however, is not peculiar to set-of-circumstances theories, and arises on any theory of information content that incorporates the conjunction of theses (*B*) and (*R*), including the structured-singular-proposition theory that I advocate. The difference is that the structured-singular-proposition view (in conjunction with (*B*) and (*R*)), unlike the theory of *Situations and Attitudes*, is not committed by its very nature to the validity of the derivation of (4*f*) from (4*b*). It is just that each step in the derivation of (4*f*) from (4*b*) is independently plausible.

IV

The puzzle of reflexives in propositional attitudes is related to a paradox that concerns quantification into belief contexts and that was discovered some time ago by Alonzo Church.[15] Unlike the former puzzle, however, Church's paradox presents a serious problem in particular for the theory of structured singular propositions.

FN:15

As a matter of historical fact, as of some appropriate date, King George IV was acquainted with Sir Walter Scott, but was doubtful whether Scott was the author of *Waverley*. We may even suppose that George IV believed at that time that Scott did not write *Waverley*. Yet, Church notes, if quantification into belief contexts is taken as meaningful in combination with the usual laws of the logic of quantification and identity, then the following is provable as a logical theorem using classical Indiscernibility of Identicals (Leibniz's Law):

(5) For every x and every y, if Georgy IV does not believe that $x \neq x$, if George IV believes that $x \neq y$, then $x \neq y$.

[15] 'A Remark concerning Quine's Paradox about Modality', Spanish translation in *Analisis Filosofico* 2. 1–2 (May–November 1982), 25–34; in English in this volume.

Mimicking the standard proof in quantified modal logic of the necessity of identity, Church remarks that although it is not certain, it was very likely true as of the same date that:

(6) For every x, George IV does not believe that $x \neq x$,

since it is very likely that George IV did not believe anything to be distinct from itself. Taking (6) as premiss, we may derive:

(7) For every x and every y, if George IV believes that $x \neq y$, then $x \neq y$.

We are thus apparently led to ascribe to King George's beliefs the strange "power to control the actual facts about x and y". Since Scott is in fact the author of *Waverley*, this derivation of (7) from (6) seems to preclude King George's believing, as of the same date, that Scott did not write *Waverley*. The derivation thus constitutes an unacceptable paradox, not unlike Russell's paradox of naïve set theory (set theory with unrestricted comprehension). Church concludes that this provides a compelling reason to reject the meaningfulness of quantification into belief contexts.[16]

FN:16

As quantification into belief contexts goes, so goes the theory of structured singular propositions as potential objects of belief. Church's paradox thus poses a serious difficulty for the theory that I advocate. But it also poses a serious difficulty for any theory, including any set-of-circumstances theory, that purports to make sense of *de re* constructions or quantification into belief contexts. Furthermore, the paradox is quite independent of the conjunction of theses (*B*) and (*R*). Whether these are true or false, the paradox arises as long as quantification into belief contexts is regarded as meaningful.

V

It is precisely to treat philosophical puzzles and problems of the sort presented here that I proposed the sketch of an analysis of the binary belief relation between believers and propositions (sometimes Russellian singular propositions) in *Frege's Puzzle*. I take the belief relation to be, in effect, the existential generalization of a ternary relation,

[16] He compares his result to the derivation in standard quantified modal logic of the contrapositive of the necessity of identity: If any x and y can be distinct, they are. He likewise cites the derivability of this principle (which he calls 'a variant of Murphy's Law') as providing a reason for rejecting the meaningfulness of quantification into modal contexts.

Church seems to allow that on the theory of structured singular propositions as potential objects of belief (which he calls 'the principle of transparency of belief' and which he regards as a doubtful theory), a power to control the actual facts about x and y with one's beliefs would not be surprising and could be explained. Unfortunately, he does not provide the alleged explanation. I am unsure what he has in mind with his remarks in this connection. Speaking as one who is deeply committed to the theory in question, I would find such a power surprising in the extreme and utterly inexplicable. I see the problem as one of how to reconcile the derivation of (7) with the obvious fact that no such power exists (and the fact that (7) may even be false in the case of Sir Walter Scott and the author of *Waverley*).

BEL, among believers, propositions, and some third type of entity. To believe a proposition *p* is to adopt an appropriate favourable attitude toward *p* when taking *p* in some relevant *way*. It is to agree to *p*, or to assent mentally to *p*, or to approve of *p*, or some such thing, when taking *p* a certain way. This is the *BEL* relation. The third relata for the *BEL* relation are something like *proposition guises*, or *modes* of acquaintance with propositions, or *ways* in which a believer may be familiar with a given proposition. Of course, to use a distinction of Kripke's, this formulation is far too vague to constitute a fully developed *theory* of belief, but it does provide a *picture* of belief that differs significantly from the sort of picture of propositional attitudes advanced by Frege or Russell, and enough can be said concerning the *BEL* relation to allow for at least the sketch of a solution to certain philosophical puzzles, including the original puzzle generated by the Richard–Soames problem.

In particular, the *BEL* relation satisfies the following three conditions:

(i) *A* believes *p* if and only if there is some *x* such that *A* is familiar with *p* by means of *x* and *BEL(A, p, x)*.

(ii) *A* may believe *p* by standing in *BEL* to *p* and some *x* by means of which *A* is familiar with *p* without standing in *BEL* to *p* and all *x* by means of which *A* is familiar with *p*.

(iii) In one sense of 'withhold belief', *A* withholds belief concerning *p* (either by disbelieving or by suspending judgement) if and only if there is some *x* by means of which *A* is familiar with *p* and not-*BEL(A, p, x)*.

These conditions generate a philosophically important distinction between withholding belief and failure to believe (i.e. not believing). In particular, one may both withhold belief from and believe the very same proposition simultaneously. (Neither withholding belief nor failure to believe is to be identified with the related notions of disbelief and suspension of judgement—which are two different ways of withholding belief, in this sense, and which may occur simultaneously with belief of the very same proposition in a single believer.)

It happens in most cases (but not all) that when a believer believes some particular proposition *p*, the relevant third relatum for the *BEL* relation is a function of the believer and some particular *sentence* of the believer's language. Consider for example the binary function *f* that assigns to any believer *A* and sentence *S* of *A*'s language, the *way A* takes the proposition contained in *S* (in *A*'s language with respect to *A*'s context at some particular time *t*) were it presented to *A* (at *t*) through the very sentence *S*. Then (assuming *t* is the time in question) Lois believes the proposition that she will inform Clark Kent of Superman's danger with her note by virtue of standing in the *BEL* relation to this proposition together with the result of applying the function *f* to Lois and the particular sentence 'I will inform Clark Kent of Superman's danger with my note.' That is, in the example the following is true:

BEL(Lois, that she will inform Clark Kent of Superman's danger with her note, *f* [Lois, 'I will inform Clark Kent of Superman's danger with my note']).

On the other hand, the following is false:

BEL(Lois, that she will inform Superman of his danger with her note, *f* [Lois, 'I will inform Superman of his danger with my note']).

Similarly, assuming the astronomer in Soames's example spoke English:

BEL(the astronomer, that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Phosphorus, *f* [the astronomer, ' "Hesperus" refers to Hesperus whereas "Phosphorus" refers to Phosphorus']),

but not:

BEL(the astronomer, that 'Hesperus' refers to Hesperus and 'Phosphorus' refers to Hesperus, f [the astronomer, ' "Hesperus" and "Phosphorus" both refer to Hesperus']).

In *Frege's Puzzle* the *BEL* relation and the function *f* are invoked in various ways to explain and to solve some of the standard (and some nonstandard) problems that arise on the sort of theory I advocate. This device is also useful with regard to the original puzzle that arises from the Richard–Soames problem and the puzzle of reflexives in propositional attitudes.

In the first example, (1*c*) is false, since Lois does not adopt an appropriate favourable attitude toward the proposition that there is someone whom she will inform of his own danger with her note, no matter how this proposition might be presented to her. That is, there is no *x* such that Lois stands in *BEL* to the proposition that she will inform someone or other of his own danger with her note and *x*. Similarly, in Soames's example. (2*c*) if false, since the astronomer does not adopt the appropriate favourable attitude toward the proposition that 'Hesperus' and 'Phosphorus' are coreferential, no matter how this proposition might be presented to him. He does not stand in *BEL* to this proposition and any *x*.

What about (1*b*) and (2*b*)? These are indeed true in the examples. Consider the first example. Sentence (1*a*) is true by hypothesis. Now notice that if Superman were somehow made aware of the truth of (1*a*), then he could truthfully utter the following sentence:

(1bI) Lois believes that she will directly inform me of my danger with her note.

In fact, (1*b*I) yields the only natural way for Superman to express (to himself) the very information that is contained in (1*a*). But if (1*b*I) is true with respect to Superman's context, then (1*b*) is true with respect to ours. Both (1*b*I), taken with respect to Superman's context, and (1*b*), taken with respect to ours, are true precisely because Lois adopts the appropriate favourable attitude toward the proposition about Superman, i.e. Clark Kent, that she will inform him of his danger with her note. Lois assents to this information when she takes it the way she would if it were presented to her through the sentence 'I will inform Clark Kent of Superman's danger with my note'. Hence, she believes it. Similarly, the astronomer inwardly assents to the proposition about Hesperus, i.e. Venus, that 'Hesperus' refers to it and 'Phosphorus' refers to it, when it is presented to him through the sentence " 'Hesperus' refers to Hesperus whereas 'Phosphorus' refers to Phosphorus". Hence (2*b*) is true.

In fact, in the examples Lois also believes that she will *not* inform Superman of his danger with her note, and the astronomer that 'Hesperus' and 'Phosphorus' do *not* both refer to Hesperus, since:

BEL(Lois, that she will not inform Superman of his danger with her note, *f* [Lois, 'I will not inform Superman of his danger with my note'])

and:

BEL(the astronomer, that 'Hesperus' and 'Phosphorus' do not both refer to Hesperus, *f* [the astronomer, ' ''Hesperus'' and ''Phosphorus'' do not both refer to Hesperus']).

Both Lois and the astronomer thus (unknowingly) believe some proposition together with its denial.[17]

FN:17

One reason so many instances of schema (L_1) are true, although it fails in these special cases, is that the schema approximates the following weaker schema, all (or at least very nearly all) of whose instances are true, and which is not falsified in these special cases:

(L_2) If $(\exists p)BEL(c, p, f[c, `\phi_a'])$, then $(\exists q)BEL(c, q, f[c,$ 'Something is such that ϕ_{it}'$])$,

where c refers to a normal speaker of English, a is any referring singular term of English, ϕ_{it} is any English sentence in which the pronoun 'it' occurs (free and not in the scope of quotation marks, an existence predicate, or other such operators) and which may also contain occurrences of a, and ϕ_a is the result of substituting (free) occurrences of a for (free) occurrences of 'it' throughout ϕ_{it}. I submit that the similarity of the former schema (L_1) to something like schema (L_2) is a major source of the plausibility of the alleged general law concerning the former. Schema (L_2) is not falsified in these special cases, even if Lois and the astronomer are normal speakers of English, since Lois does not agree to the proposition that she will inform Superman of his danger with her note when she takes it in the way she would if it were presented to her through the sentence 'I will inform Superman of his danger with my note', and the astronomer does not agree to the proposition that 'Hesperus' and 'Phosphorus' refer to Venus when it is presented to him through the sentence ' ''Hesperus'' and ''Phosphorus'' refer to Hesperus.'[18]

FN:18

[17] That is, neither Lois nor the astronomer knows (in the example) that she or he believes some proposition together with its denial. On the other hand, Lois does know that she believes that she will inform Superman of his danger with her note and will not inform Superman of his danger with her note, and the astronomer does know that he believes that 'Hesperus' refers to Venus and 'Phosphorus' refers to Venus and 'Hesperus' and 'Phosphorus' do not both refer to Venus. Furthermore, each presumably knows that these propositions are contradictory—though neither knows that she or he believes a contradictory proposition. Sorting these matters out is a delicate task made extremely difficult by relying on the term 'believes' without the use of some expression for the full ternary *BEL* relation. Cf. *Frege's Puzzle*, ch. 8.

[18] Soames has offered an account not unlike this one in response to my urging on him that, as Richard originally presented the problem, it poses a serious difficulty for the theory of structured singular propositions as well as for the set-of-circumstances theories. See 'Lost Innocence', p. 69 n. 12. The notion of a ''belief state'' invoked there (which seems to have been derived from 'The Problem of the Essential Indexical' and *Situations and Attitudes*) plays a role analogous to the third relata of the *BEL* relation in my account, the *ways* in which one may be familiar with, or take, a

VI

Even if this resolves the original puzzle generated by the Richard–Soames problem for the structured-singular-proposition account of information content, it does not yet lay to rest the puzzle of reflexives in propositional attitudes, not to mention Church's ingenious paradox concerning quantification into belief contexts.

Richard's proposal to solve the original puzzle by blocking the initial inference from (3*a*) (together with the fact that *a* and *b* are co-referential proper names or other simple singular terms) to (3*b*) would equally block the puzzle of reflexives in propositional attitudes. This proposal involves relinquishing thesis (*B*), and is motivated by the threat of the alleged derivability of falsehoods such as (1*c*) from (1*b*). But I argued above that thesis (*B*) is supported by strong linguistic evidence, and should be maintained insofar as the facts allow. We have seen that the account of belief in terms of the *BEL* relation effectively blocks the move from (1*b*) to (1*c*), while retaining thesis (*B*) and while also affording an explanation (or at least the sketch of an explanation) for the *prima facie* plausiblity of the move. If there is a solution to the problem of reflexives in propositional attitudes, it does not lie in the rejection of thesis (*B*).

Ruth Barcan Marcus has argued that, in at least one ordinary sense of 'believe', it is impossible to believe what is impossible.[19] Marcus would thus claim that (4*a*) is false to begin with, since the astronomer cannot "enter into the belief relation" to the information, which is necessarily misinformation, that Hesperus outweighs

FN:19

proposition. See *Frege's Puzzle*, ch. 9 n. 1, for some brief remarks comparing the third relata of the *BEL* relation (proposition guises, or modes of acquaintance with propositions) with Perry's notion of a belief state.

In response to the Richard–Soames problem, Barwise and Perry seem to have abandoned the idea that the information content ("interpretation") of a declarative sentence, with respect to a context, is the set of situations (or type of situation) with respect to which the sentence is true (with 'situation' understood in such a way that any situation with respct to which it is true that 'Hesperus' and 'Phosphorus' both refer to Venus is one with respect to which it is also true that there is something referred to by both 'Hesperus' and 'Phosphorus'). In fact, Barwise and Perry seem to have moved significantly in the direction of structured singular propositions and an account of the Richard–Soames problem similar (in certain respects) to the one advanced here and to the one advanced by Soames (see 'Shifting Situations and Shaken Attitudes', *Linguistics and Philosophy* 8 (1985), 105–161 (also available as the Stanford University Center for the Study of Language and Information Report No. CSLI-84-13), pp. 153–158, esp. pp. 156–157). If so, this move constitutes an important concession to Soames. However, Barwise and Perry (if I understand them correctly) couple this move with the surprising claim (p. 158) that there is a significant sense in which the information content of "Something or other is referred to by both 'Hesperus' and 'Phosphorus' " is not a consequence of that of "Venus exists and 'Hesperus' refers to it and 'Phosphorus' refers to it". If this is to be understood as a claim about the logic of the information contents of these sentences, surely the claim must be rejected, and the doctrine of structured singular propositions as the information contents ("interpretations") of sentences and the objects of belief, coupled with classical logic, is much to be preferred over the newer theory of Barwise and Perry.

[19] 'A Proposed Solution to a Puzzle about Belief', in P. French. T. Uehling, and H. Wettstein (eds.) *Midwest Studies in Philosophy, vi. The Foundations of Analytic Philosophy*, (University of Minnesota Press, 1981), 501–510, esp. 503–506, and 'Rationality and Believing the Impossible', *Journal of Philosophy* 80, 6 (June 1983), 321–338.

Phosphorus. However, one of Marcus's arguments for this, perhaps her main argument, appears to be that where *a* and *b* are co-referential names, if (3*a*) is true so is (3*c*), and in a great many cases where one is inclined to hold an instance of (3*a*) true even through ⌜*aRb*⌝ encodes necessarily false information ((4*a*) for example), (3*c*) is patently false, because ⌜(∃*x*)*xRx*⌝ (e.g. 'Something outweighs itself') encodes information that is not only impossible but patently unbelievable.[20]

FN:20

Marcus's view that one cannot believe what cannot be true is highly implausible, and I believe, idiosyncratic. It often happens in mathematics and logic that owing to some fallacious argument, one comes to embrace a fully grasped proposition that is in fact provably false. Sometimes this happens even in philosophy, more often than we care to admit. In our example, we may suppose that, for some particular number *n*, the astronomer comes to believe the proposition that Hesperus weighs at least *n* tons, and also the proposition that Phosphorus weighs no more than $(n - 1,000)$ tons. He embraces these two propositions. It is very implausible to suppose that the fact that their conjunction is such that it could not be true somehow prevents the astronomer from embracing that conjunction, along with its component conjuncts, or that the astronomer is somehow prevented from forming beliefs on the basis of inference from his two beliefs, as in (4*a*).

More important for our present purpose is that Marcus's argument for the falsehood of (4*a*), at least as the argument is interpreted here, has to be mistaken. Otherwise, one could also show that (1*a*) and (2*a*) are false in the original examples. For although the proposition that Lois will inform someone of his own danger with her note is not unbelievable, it is plain in the example that it is not believed by Lois, i.e. (1*c*) is plainly false. If Marcus's argument for the impossibility of believing the impossible were sound, then by parity of reasoning it would follow that (1*a*) is false. Similarly, although the proposition that 'Hesperus' and 'Phosphorus' are co-referential is believable, in fact true, it is a hypothesis of the example that the astronomer does not believe it, i.e. (2*c*) is stipulated to be false. If Marcus's claim that (3*c*) is true if (3*a*) is true were itself true, it would follow that (2*a*) is false. But (1*a*) and (2*a*) are plainly true in these examples. There must be something wrong, therefore, with Marcus's argument, at least as I have interpreted it here.

What is wrong is precisely the claim that (3*c*) is true if (3*a*) is. Since it is incorrect, this claim cannot give us a way out of the present problem. In fact by shifting from (4*a*)–(4*f*) to another example, we can remove the feature that what is said to be believed at step (*a*) is such that it could not be true. Thus from 'Lois believes that

[20] See 'A Proposed Solution to a Puzzle about Belief', p. 505, and 'Rationality and Believing the Impossible', p. 330. Marcus's argument focuses on the special case where *R* is a predicate for numerical distinctness. She writes: "If I had believed that Tully is not identical with Cicero, I would have been believing that something is not the same as itself and I surely did not believe that, a blatant impossibility, so I was mistaken in claiming to *have* the belief [that Tully is not Cicero]", and, "[believing that London is different from Londres] would be tantamount to believing that something was not the same as itself, and surely I could never believe *that*. So my belief claim [my claim that I believed that London is not Londres] was mistaken . . .". These arguments evidently rely on the premiss that if (3*a*) then (3*c*) (or perhaps on the premiss that if (3*a*) then [the existential generalization on *a* of (3*e*′)], i.e. if *c* believes that *a* and *b* are distinct, then *c* believes *of* something that it is distinct from itself).

Clark Kent disparages Superman while Superman indulges Clark Kent' we may construct a parallel and equally fallacious derivation of 'Lois believes that Superman is self-disparaging and self-indulgent'. Marcus's unusual contention that it is impossible to believe the impossible, whether correct or incorrect, is simply irrelevant to this example.

What, then, is the solution to the puzzle of reflexives in propositional attitudes for the theory of structured singular propositions?

In the example, (4*b*) and (4*b*′) are true, whereas (4*f*) and (4*f*′) are false. Any temptation to infer (4*f*) from (4*b*), or (4*f*′)(4*b*′), can be explained using the *BEL* relation and the function *f* in a manner similar to the explanation given above in connection with the *prima facie* plausibility of inferring (3*c*) from (3*b*). In any case, either the inference from (4*b*) to (4*e*) (and therewith the inference from (4*b*′) to (4*e*′)) is fallacious, or the inference from (4*e*) to (4*f*) (and therewith the inference from (4*e*′) to (4*f*′)) is. Which is it?

Answering this question involves taking sides in a current controversy concerning the identity or distinctness of propositions of the form *x bears R to x* and *x bears R to itself*. If the propositions that Hesperus outweighs Hesperus and that Hesperus outweighs itself are the very same, then the inference from (4*b*) to (4*e*) is valid by classical Indiscernibility of Identicals (or Leibniz's Law) together with thesis (*B*), and the inference from (4*e*) to (4*f*) must then be rejected. If, on the other hand, these propositions are not the same and instead the proposition that Hesperus outweighs itself is the same (or very nearly the same) as the proposition that Hesperus is self-outweighing, then the inference from (4*e*) to (4*f*) is unobjectionable and the inference from (4*b*) to (4*e*) must be rejected.

As noted in Section III above, the advocate of a set-of-circumstances theory of information content is committed to the claim that propositions of the form *x bears R to x* and *x bears R to itself* are exactly the same, since any circumstance in which *x* bears *R* to *x* is one in which *x* bears *R* to itself, and vice versa. Thus, M. J. Cresswell, a set-of-possible-worlds theorist, has recently claimed that:[21]

FN:21

> on any reasonable account of propositions, the proposition that Ortcutt loves himself ought to be the same as the proposition that Ortcutt loves Ortcutt.

This, however, is far from the truth. In fact, there are compelling reasons to distinguish a proposition of the form *x bears R to x* from the proposition *x bears R to itself*. One sort of consideration is the following: we must distinguish between the reflexive property of exceeding oneself in weight and the simple relational property of exceeding the planet Venus in weight. The former is an impossible property; it is quite impossible for anything to possess it. The latter property, on the other hand, is fairly widespread; a great many massive objects (e.g. the stars) possess it—although, of course, it is quite impossible for Venus to possess it. Now the sentence 'Hesperus outweighs itself' seems to ascribe to Hesperus, i.e. Venus, the impossible property of weighing more than oneself, rather than the simple relational property of weighing more than Venus. It seems to say about Venus what 'Mars outweighs itself' says about

[21] *Structured Meanings: The Semantics of Propositional Attitudes* (MIT Press, 1985), p. 23.

Mars—that it has the reflexive property of exceeding oneself in weight—and not what 'Mars outweighs Venus' says about Mars. If one wants to ascribe to Venus the simple relational property of weighing more than Venus, rather than the impossible property of weighing more than oneself, one may use the sentence 'Hesperus outweighs Hesperus' (among others). It says about Venus what 'Mars outweighs Venus' says about Mars—that it weighs more than Venus—instead of what 'Mars outweighs itself' seems to say about Mars. If one prefers, it ascribes the relation of exceeding-in-weight to the ordered pair of Venus and itself. In either case, the proposition contained in 'Hesperus outweighs Hesperus' is not the same as what seems to be the proposition contained in 'Hesperus outweighs itself'.[22] Contrary to any set-of-circumstances account of propositions, the proposition about Venus, that it weighs more than it, is a different proposition from the proposition about Venus that it is self-outweighing, although they are, in some sense, logically equivalent to one another.[23]

FN:22

FN:23

[22] The argument presented thus far has been emphasized by D. Wiggins in a number of writings. See e.g. 'Identity, Necessity and Physicalism', in S. Körner (ed.), *Philosophy of Logic* (University of California Press, 1976), 96–132, 159–82, esp. 164–6; and 'Frege's Problem of the Morning Star and the Evening Star', in M. Schirn (ed.), *Studies on Frege*, ii. *Logic and the Philosophy of Language* (Bad Canstatt: Stuttgart, 1976), 221–255, esp. 230–231.

Wiggins credits the argument to Peter Geach, and claims to have extracted the argument from Geach *Reference and Generality* (Cornell University Press, 1962), p. 132. This, however, is a serious misinterpretation of Geach, whose view is precisely the denial of Wiggins's view that sentences of the form ⌜*a* bears *R* to *a*⌝ and ⌜*a* bears *R* to itself⌝ contain different information, or express different propositions. See e.g. Geach 'Logical Procedures and the Identity of Expressions', in id. *Logic Matters* (University of California Press, 1972), 108–115, esp. 112–113. If I read Geach correctly, his view is that a sentence such as 'Hesperus outweighs Hesperus' ascribes to Venus the reflexive property of weighing more than oneself, as does the sentence 'Hesperus outweighs itself', rather than the simple relational property of weighing more than Venus. (Cf. the treatment of the contents of sentences with recurring expressions in H. Putnam, 'Synonymy, and the Analysis of Belief Sentences', *Analysis* 14, 5 (April 1954), 114–122; also in this volume. See also *Frege's Puzzle*, pp. 164–165 n. 4.) The argument in *Reference and Generality* is intended to show not that 'Marx contradicts himself' differs in information content from 'Marx contradicts Marx' (which Geach rejects), but that the 'himself' in 'Marx contradicts himself' is not a singular term referring to Marx. The argument for this conclusion (which Wiggins presumably also believes) is part of a defence of Geach's general view that pronouns occurring with antecedents are typically not referring singular terms. (I disagree with Geach both concerning the semantic analysis of sentences such as 'Hesperus outweighs Hesperus' and 'Marx contradicts Marx', as does Wiggins, and concerning typical (nonreflexive) pronouns with antecedents, though the latter issue is not germane to the topic of the present discussion.)

The same (or very nearly the same) misinterpretation of Geach's argument in *Reference and Generality* occurs in G. Evans 'Pronouns, Quantifiers and Relative Clauses (I)', in M. Platts (ed.), *Reference, Truth and Reality*, (Routledge and Kegan Paul: London, 1980), 255–317, esp. 267–268, although Evans admits that the view he attributes to Geach has been unambiguously denied by Geach in a number of places. (Oddly, Evans cites references to, and even quotes, writings in which Geach clearly denies the view that Evans attributes to him.) It might be said that in accusing Geach of misinterpreting Geach, Evans takes on the very property he attributes to Geach—although Evans does not misinterpret himself.

[23] The response to the Richard–Soames problem in 'Shifting Situations and Shaken Attitudes' suggests that Barwise and Perry might similarly respond to the puzzle of reflexives in propositional attitudes by claiming that 'Hesperus outweighs Hesperus' and 'Hesperus is self-outweighing' differ in information content (have different "interpretations"). See n. 18 above. Such a move would constitute a repudiation of the idea, fundamental to *Situations and Attitudes*, that the information content ("interpretation") of a sentence is the set of situations (or type of situation) with respect to

The astronomer in the example believes the former and not the latter. Neither the sentence 'Hesperus outweighs Hesperus' nor the sentence 'Hesperus outweighs itself' can be regarded as somehow containing *both* of these propositions simultaneously (as might be said, for example, of the conjunction 'Venus has the simple relational property of weighing more than Venus and also the reflexive property of weighing more than oneself'). Each sentence contains precisely one piece of information, not two. Neither is ambiguous; neither is a conjunction of two sentences with different (albeit equivalent) information contents.[24] Similar remarks may be made in connection with Cresswell's example of 'Ortcutt loves Ortcutt' and 'Ortcutt loves himself'.

FN:24

This conception of reflexive propositions of the form *x bears R to itself* involves rejecting the otherwise plausible view that the reflexive pronoun 'itself' in 'Hesperus outweighs itself' refers anaphorically to the planet Venus. Instead, the pronoun might be regarded as a predicate-operator, one that attaches to a dyadic predicate to form a compound monadic predicate. Formally, this operator may be defined by the following expression:[25]

FN:25

$$(\lambda R)(\lambda x)xRx.$$

The alternative conception of propositions of the form *x bears R to itself* involves treating reflexive pronouns instead as anaphorically referring singular terms. On this view, in order to ascribe to Venus the reflexive impossible property of weighing more than oneself, it is not sufficient to use the sentence 'Hesperus outweighs itself'. Instead, one must resort to some device such as the predicate 'is self-outweighing'.

There can be no serious question about the possibility of an operator such as the one defined above. The displayed expression definitely captures a *possible* operator on dyadic predicates. There is no reason why English (and other natural languages) could not contain such an operator, and there is no *a priori* argument that standard English does not have this operator. The question is whether the reflexive pronouns

which the sentence is true—with the term 'situation' understood in such a way that any situation with respect to which it is true that Venus outweighs Venus is one with respect to which it is also true that Venus is self-outweighing, and vice versa. Any attempt to modify their view to accommodate the fact that 'Hesperus outweighs Hesperus' and 'Hesperus is self-outweighing' differ in information content would clearly constitute a concession to the structured-singular-proposition sort of theory advocated in *Frege's Puzzle*.

However, this move might be coupled with the claim that there is a significant sense in which the second information content (''interpretation'') is no consequence of the first. They might even claim that there is a significant sense in which these information contents are independent and neither implies the other. Here again, if either of these claims is to be understood as concerning the logic of the information contents of 'Hesperus outweighs Hesperus' and 'Hesperus is self-outweighing', they must surely be rejected. Insofar as the newer theory of Barwise and Perry includes one or both of these claims, the doctrine of structured singular propositions coupled with the denial of each of these claims (and with classical logic) is much the preferable theory.

[24] This part of the argument is intended as a rejoinder to Evans's response in 'Pronouns, Quantifiers and Relative Clauses', p. 268.

[25] Cf. *Reference and Generality*, pp. 136–137. In the theoretical apparatus of *Frege's Puzzle*, the contribution to information content made by (i.e. the ''information value'' of) the displayed expression is the *operation* of assigning to any class K of ordered pairs of individuals the class of individuals i such that the reflexive pair $\langle i, i\rangle \in K$.

of standard English ('itself', 'himself', 'myself', 'oneself', etc.) are expressions for this operator, rather than anaphorically referring singular terms.

This is not a metaphysical question about the essential natures of propositions, but an empirical question about the accidents of standard English semantics. It is a question, moreover, for which decisive linguistic evidence is difficult to produce, since on either hypothesis the information content of 'Hesperus outweighs itself' is logically equivalent to the content yielded by the rival hypothesis (although writers on both sides of this dispute have advanced what they take to be compelling evidence for their view).

Assuming that the semantic analysis presented above of sentences such as 'Hesperus outweighs Hesperus' is at least roughly correct, the claim that propositions of the form *x bears R to x* and *x bears R to itself* are the same is tantamount to the empirical claim that the reflexive pronouns of standard English are singular terms and not expressions for the predicate-operator defined above, whereas the claim that the proposition *x bears R to itself* is not the same as *x bears R to x* but instead goes with *x is self-R* is tantamount to the empirical claim that the reflexive pronouns are expressions for the predicate-operator and not singular terms. This issue cannot be settled by *a priori* philosophical theorizing about the nature of propositions. A complete solution to the puzzle of reflexives in propositional attitudes thus turns on answering a difficult empirical question concerning the meanings of reflexive pronouns in standard English.

VII

The time has come to face the music. How can the theory of structured singular propositions solve Church's paradox concerning quantification into belief contexts?

Fortunately, some of the ideas discussed in the preceding sections bear directly on Church's paradox. Notice first that (7), taken literally, does not ascribe any power to King George or his beliefs *per se*. Nor does it ascribe to George an infallibility concerning the distinctness of distinct individuals *x* and *y*. It merely states a generalization concerning every pair of individuals *x* and *y* believed distinct by King George. In Humean terminology, it merely states a *constant conjunction* between any pair of individuals being believed distinct by King George and their actually being distinct. As Hume noted, there is no idea of power contained in that of constant conjunction. Analogously, the sentence 'All crows are black' merely states a generalization, or constant conjunction, concerning all crows. The idea that something's being a crow somehow *makes it* black arises only when this sentence is regarded as having the status of biological law, rather than that of a purely accidental generalization.

Likewise, the conclusion (7) can be regarded as ascribing a power or nomological regularity to King George's beliefs only if (7) is regarded as having the status of a law ascribing some special law-governed feature to George IV and his beliefs, rather than as an accidental constant conjunction. Now in deriving (7), we took (6) as our only premiss. Thus (7) may be regarded as stating some sort of law only if (6) may be.

Church remarks that, even though (6) is not certain, it is very likely. This observation may support a plausible view of (6) as some sort of psychological law concerning George IV and his beliefs. In this way, (7) would emerge as a law ascribing a nomological feature to King George's beliefs. Since no such law in fact obtains, and may even be falsified by the very case of Sir Walter Scott and the author of *Waverley*, the meaningfulness of quantification into belief contexts, and therewith the theory of structured singular propositions, would be thereby discredited.

On the theory that I advocate, however, (6) is not only not very likely, as of some particular date during King George's acquaintance with Scott, it is very likely false.

It may seem as if denying (6) is tantamount to saying that George IV believed of some x that it is distinct from itself, and this seems a serious charge indeed. If an interest in the law of identity can hardly be attributed to the first gentleman of Europe, it is nothing short of blasphemy to attribute to him an interest in denying that law. In claiming that (6) is very likely false, as of some appropriate date, I mean no disrespect. Sentence (6) can easily be false even though King George is, of course, entirely rational—in fact, even if he were (what is beneath his dignity) a master of classical logic. If there was some time when George IV was acquainted with Scott and nevertheless believed after reading a *Waverley* novel that Scott was not the author, then (6) is false with respect to that time. If this be disputed, imagine instead that George IV confronted Scott at a book-signing ceremony, at which Scott truthfully proclaimed his authorship of *Waverley* but disguised himself in order to conceal his identity as Sir Walter Scott. Suppose the disguise succeeded in fooling even King George.[26] Let George IV say with conviction, pointing to the disguised author, 'He is not Sir Walter Scott'. In this case, (6) is decisively false. George IV is in the same unfortunate position as that of the ancient astronomer who believed of Venus that it is distinct from it.

FN:26

Why, then, does Church claim that (6) is very likely? My conjecture is that Church confuses (6) with:

(6′) For every x, George IV does not believe that $(\lambda x')[x' \neq x']\ (x)$

or with:

(6″) For every x, George IV does not believe that x is self-distinct,

where the term 'self-distinct' is understood in accordance with the definition of the '*self*'-prefix given in Section III above. Both of these are indeed extremely likely—nay (I hasten to add), virtually certain. On the theory that I advocate, the pair of open sentences

$$x \neq x$$

and

$$(\lambda x')[x' \neq x']\ (x)$$

[26] As a matter of historical fact, Scott did conceal his authorship of *Waverley*, and George IV did wish to know whether Scott indeed wrote *Waverley*. Hence Russell's clever example.

(or 'x is distinct from x' and 'x is self-distinct'), although logically equivalent, must be sharply distinguished as regards the propositions expressed under any particular assignment of a value to the variable 'x'. Under the assignment of Scott to 'x', the singular proposition contained in the first open sentence is believed by George IV in the book-signing example, the second is not. The extreme likelihood of (6′) and (6″) does not extend to (6).

Whereas sentences (6′) and (6″) are similar to, and easily confused with sentence (6), the former sentences do not concern King George's doxastic attitudes toward the propositions involved in sentence (6). They concern propositions of the form *x is self-distinct* (which ascribe the plainly impossible property of self-distinctness to particular individuals x) rather than propositions of the form *x is distinct from x* (which ascribe the relation of distinctness to reflexive pairs of individuals $\langle x, x\rangle$). Sentences (6′) and (6″) provide adequate explanation why George IV is disinclined to answer affirmatively when queried "Is Sir Walter self-distinct?", but the substitution of these sentences for Church's (6) does not show sufficient appreciation for the fact that King George is similarly disinclined when queried "Is Sir Walter distinct from Sir Walter?", or when any other similarly worded question is posed. These considerations give rise to a second potential confusion that could also lead one to conclude erroneously that (6) is true or at least very likely. By invoking the ternary *BEL* relation, something even closer to (6) may be assumed as at least very likely:

(6″′) For every x, if there is a y such that George IV is familiar with the proposition that $x \neq x$ by means of y, then there is a y' such that George IV is familiar with the proposition that $x \neq x$ by means of y' and not-*BEL*(George IV, that $x \neq x$, y').

That is, either George IV is not familiar at all with the proposition that $x \neq x$ (in which case he does not believe it) or he withholds belief concerning whether $x \neq x$, either by disbelieving or by suspending judgement. (See the third condition on the *BEL* relation in Section V above.) Although (6‴) is not certain, it is very likely true as of the date in question, and this yields an explanation for King George's failure to assent to 'Sir Walter is distinct from Sir Walter'. But (by the first and second conditions on *BEL*) it does not follow that (6) itself is true or even likely.

It is entirely an empirical question whether (6) itself is true. There is no reason in advance of an actual investigation to suppose that (6) is even probably true.[27] By the same token, however, even if (6) is in fact very unlikely, it might well have been true throughout King George's lifetime. In some perfectly plausible alternative history of the world, it is true. If (6) were true, (7) would be as well. What then? Are we only contingently rescued from paradox in the actual world by the contingent falsity of (6)?

FN:27

Even if (7) were true, it would not state a law ascribing some strange property to King George's beliefs. It would state a purely contingent constant conjunction concerning every pair of individuals x and y, an accidental generalization that happens to

[27] This contrasts sharply with the analogous principle involved in the standard proof in quantified modal logic of the necessity of identity: For every x, it is not possible that $x \neq x$. This is a logical truth, and therefore an *a priori* certainty. Unlike (7), the necessity of identity (or equivalently, Murphy's Law of Modality) is a genuine law (in this case, a law of logic).

be true not by virtue of some nomological feature of King George IV and his beliefs, but because—fortunately for King George—(6) happens to be true. No power to control the actual facts about x and y would be ascribed to King George's beliefs. If (6) were true (and Scott still had written *Waverley*), it would have to be true as well that King George does not believe that Scott is not the author of *Waverley*, and that George IV is not otherwise mistaken about the distinctness of any other pairs of identical objects of his acquaintance. The derivation of (7) from (6) would be sound, but it would no more constitute an unacceptable paradox than the so-called "paradoxes of material implication" constitute unacceptable paradoxes concerning 'if . . ., then'. In fact, since (7) employs the material 'if . . ., then', Church's paradox concerning quantification into belief contexts is a version of one of the "paradoxes of material implication".

VIII

What is the nature of the connection among the Richard–Soames problem, the puzzle of reflexives in propositional attitudes, and Church's "paradox" concerning quantification into belief attributions?

It is important to notice that, unlike the original puzzle generated by the Richard–Soames problem, neither the puzzle of reflexives in propositional attitudes nor Church's paradox makes essential use of existentially general beliefs, such as those ascribed in (1*c*), (2*c*), or (4*c*), or that denied in:

George IV does not believe that for some x, $x \neq x$.

Instead, the puzzle of reflexives in propositional attitudes and Church's paradox essentially employ beliefs whose formulation involves reflexive devices, such as the reflexive pronoun 'itself' and the '*self*'-prefix defined above. Conversely, the original puzzle, as constructed by means of sentences such as (1*b*), (2*b*), and (4*b*), makes no explicit use of beliefs whose formulations involve reflexive pronouns or other such devices. In lieu of such beliefs, the original puzzle employs beliefs whose formulations involve repeated occurrences of the same, or otherwise anaphorically related, bound variables or pronouns: the occurrences of 'x' in (3*c*), the occurrences of 'it' in (2*c*) and (4*c*), the 'whom' and 'his' in (1*c*). In each case, these recurrences, or similarly related occurrences, are bound together from *within* the belief context. If I am correct, Church's "paradox" results, in part, from a confusion of a belief involving recurrences of the same variable bound together from *outside* the belief context with a belief involving a reflexive device. Nothing with the force of any of these puzzles is generated if we confine ourselves to beliefs involving recurrences of the same proper name, as in (1*b*), (2*b*), and (4*b*), or beliefs involving recurrences of the same variable or pronoun bound together from without, as ascribed in (3*d*) and (4*b*′) and denied in (6), and keep them sharply separated from beliefs involving reflexive devices or variables or pronouns bound together from within. On the theory formed from the conjunction of theses (*B*) and (*R*), sentences (1*b*), (2*b*), (4*b*), and (4*b*′) are all straightforwardly true. It appears likely, therefore, that the general phenomenon that

gives rise to all three of these puzzles centres on some important element that is common to beliefs whose formulations involve reflexive devices and beliefs whose formulations involve recurrences of variables or pronouns bound together (from within any belief attribution), but absent from beliefs whose formulations involve recurrences of proper names or of free variables or pronouns (bound together from without the belief attribution).

Wherein is this common element of reflexivity? The question is significantly vague, and therefore difficult to answer. Some of the apparatus of *Frege's Puzzle*, however, points the way to a possible response.

In *Frege's Puzzle* the binding of a variable is regarded as involving the abstraction of a compound monadic predicate from an open sentence. Thus '($\exists x$)('Hesperus' refers to x and 'Phosphorus' refers to x)' is seen on the model of 'Something is such that 'Hesperus' refers to it and 'Phosphorus' refers to it', and '($\exists x$)(x outweighs x)' is seen on the model of 'Something is such that it outweighs it', where in each case the initial word 'something' is a second order predicate and the remainder of the sentence is the abstracted compound monadic predicate to which 'something' is attached. In fact, the abstracting of a predicate from an open sentence of formal logic using Church's 'λ'-operator might be understood on the model of transforming an "open" sentence such as 'I love *it* and *it* loves me' (with both occurrences of '*it*' functioning as "freely" as a free variable of formal logic) into the corresponding closed monadic predicate 'is such that I love it and it loves me'.

Compound monadic predicates formed by variable-binding (or pronoun-binding) abstraction from open sentences are treated in *Frege's Puzzle* as yielding an exception to the general rule that the contribution to information content made by (i.e. the "information value" of) a compound expression is a complex entity made up of the contributions of the components. Instead such compound predicates are taken as contributing a semantically associated temporally indexed property, taken as a unit. (See note 5.) Thus, the (closed) abstracted predicate 'is an object x such that 'Hesperus' refers to x and 'Phosphorus' refers to x' is regarded as contributing, with respect to a time t, simply the property of being referred to at t by both 'Hesperus' and 'Phosphorus', and the (closed) abstracted predicate 'is an object x such that x outweighs x' is regarded as contributing, with respect to t, the property of outweighing oneself at t. The proposition contained, with respect to t, by 'Something is such that it outweighs it' (or 'Something is an object x such that x outweighs x') is taken as being composed of this latter property together with the contribution made by 'something' (to wit, the property of being a nonempty class at t).

The properties of being referred to at t by 'Hesperus' and also by 'Phosphorus' and of outweighing onself at t contain the element of reflexivity that also arises when using the '*self-*' prefix, defined in Section III above by means of the binding of a recurring variable. The dyadic-predicate-operator defined in Section VI above in connection with the question of the meanings of reflexive pronouns also involves the binding of a recurring variable, and thereby also involves this element of reflexivity. Some such aspect of the binding of recurring variables and pronouns seems to provide the link among the Richard–Soames problem, the puzzle of reflexives in propositional attitudes, and Church's paradox concerning quantification into belief contexts.