

A NEW PUZZLE ABOUT INTENTIONAL IDENTITY*

I. HISTORICAL BACKGROUND: GEACH'S PUZZLE

In 1967 Peter Geach [7] introduced a new problem for the semantics of statements containing psychological verbs. Sentences like (1) below generate the puzzle.¹

- (1) Hob thinks a witch blighted Bob's mare, and Nob
thinks she killed Cob's sow.

Because such statements appear to assert an identity between two objects of thought, Geach refers to them as statements of "intentional identity". Uses of an intentional identity statement such as (1) may differ along various dimensions, and before I explain just how Geach's puzzle arises it will help to distinguish at least one of these dimensions at an intuitive level. *Some* uses of (1) commit the speaker to the existence of a witch, or at least to the existence of a real object that Hob (and possibly Nob) take to be a witch. *Other* uses of (1) do not commit the speaker to the existence of anything Hob's and Nob's beliefs are about. At this point I am speaking only about linguistic intuitions regarding possible uses of the sentence; philosophical questions about what sense can be made of these intuitions are not yet being raised. That certain uses of (1) commit the speaker to the existence of an object on which Hob's and Nob's beliefs are focused is clear enough, but it might be helpful to bolster intuitions about the other kind of use by considering two situations in which a use of (1) would be true even though no existing object is the common focus on Hob's and Nob's beliefs.

EXAMPLE 1. Last night, Bob's mare became quite ill. Hob, who tends Bob's barn, inferred that a witch blighted her. This morning Hob said to his friend, Nob, "A witch blighted Bob's mare."

Nob believes what Hob has told him. He thinks for a moment, and says, “Cob’s sow died early this morning. I’ll bet the same witch killed the sow, too.” But in fact both animals fell ill due to perfectly natural causes.

EXAMPLE 2. The Gotham City newspapers have reported that a witch, referred to as “Samantha”, has been on quite a rampage. According to the article she has been blighting farm animals and crops and throwing people down wells. In reality, there is no such person: the animals and crops all died of natural causes, and the people found at the well-bottoms had all stumbled in by accident in a drunken stupor. The news reporters simply assumed that a witch was responsible for all the mishaps, and dubbed her “Samantha”. Hob and Nob both read the *Gotham Star* and, like most folks, they believe the stories about the witch. Hob thinks Samantha must have blighted Bob’s mare, which took ill yesterday. Nob thinks Samantha killed his friend Cob’s sow. (For purposes of later discussion, we assume Nob has no beliefs at all about Hob or about Bob’s mare; he is unaware of the existence of either.)

It will be useful to settle on some terminology for referring to the two kinds of uses of intentional identity statements. If we use existential import as the criterion of whether a construction is being used de re or de dicto, we can say that the first kind of use is de re, and the second de dicto, with respect to the noun phrase ‘a witch’ and the two occurrences of ‘thinks’.²

Geach’s puzzle arises when one inquires about the logical form of (the two uses of) (1); in particular when one inquires how the anaphoric relation between the pronoun ‘she’ and its antecedent ‘a witch’ is to be understood. Suppose you say the pronoun is to be understood *quantificationally*, as a variable bound by its antecedent quantifier phrase. This is largely a syntactic hypothesis.³ The hypothesis is that (1) can be formalized in fairly standard formal languages as (2), where ‘ \mathbf{B}_α ’ formalizes ‘ α believes that’.

$$(2) \quad (Ex)[Wx \ \& \ \mathbf{B}_h(Bx) \ \& \ \mathbf{B}_n(Cx)].$$

Assume that the quantifier in (2) carries existential import. That is, assume that (2) counts as an adequate formalization of (3) below (and that appropriate assignments of representational role have been made to relevant features of the algebraic structures used to interpret the formal language, in a way consistent with such a formalization scheme).

- (3) There exists a witch such that Hob believes she blighted Bob's mare, and such that Nob believes she killed Cob's sow.

In that case (2) will apparently serve as an adequate formalization of the de re uses of (1), for such uses are true if and only if (3) is true. But then clearly (2) is an unacceptable formalization of the de dicto uses of (1): such uses of (1) can be true even when (3) is false. So if the quantifier in (2) carries existential import, it cannot serve as a formalization of de dicto uses of (1).

An alternative hypothesis is that the pronoun in a de dicto use of (1) is not a quantificational pronoun but what Geach calls a *pronoun of laziness*. This is a pronoun that simply goes proxy for some expression constructible from words occurring in the syntactic environment of its antecedent, and which is employed primarily to avoid repetitious language.⁴ Here, the suggestion is that in de dicto uses of (1), the pronoun goes proxy for a definite description such as 'the witch who blighted Bob's mare' or 'the witch Hob thinks blighted Bob's mare'. On this hypothesis, de dicto uses of (1) would abbreviate either (4) or (5), in which the definite description takes narrowest possible scope.

- (4) Hob thinks a witch blighted Bob's mare, and Nob thinks the witch who blighted Bob's mare killed Cob's sow.
- (5) Hob thinks a witch blighted Bob's mare, and Nob thinks the witch Hob thinks blighted Bob's mare killed Cob's sow.

Since, on this hypothesis, de dicto uses of (1) abbreviate either (4) or (5), such uses of (1) would be formalized just as (4) or (5) are; that is, as (6) or (7).

- (6) $\mathbf{B}_h(Ex)(Wx \ \& \ Bx) \ \& \ \mathbf{B}_h[(Ex)(Wx \ \& \ Bx)$
 $\ \& \ (y)(Wy \ \& \ By \ \rightarrow \ x = y) \ \& \ Cx]$
- (7) $\mathbf{B}_h(Ex)(Wx \ \& \ Bx) \ \& \ \mathbf{B}_h\{(Ex)[\mathbf{B}_h(Wx \ \& \ Bx)$
 $\ \& \ (y)(\mathbf{B}_h(Wy \ \& \ By) \ \rightarrow \ x = y) \ \& \ Cx]\}$.

Initially the suggestion seems reasonable enough. Certainly the intended readings of (4) and (5) are true in the situation described in Example 1. But it is easy to see that the suggestion will not work. Although de dicto uses of (1) are true in Example 2, neither (4) nor (5) is true in Example 2: Nob knows nothing about Hob or about Bob's mare. The pronouns in at least some de dicto intentional identity statements are not pronouns of laziness.

Geach's puzzle, then, can be put in the following way. The pronouns in de dicto intentional identity statements cannot be interpreted quantificationally if existential quantifiers retain their existential import. Nor can they be interpreted as pronouns of laziness. How, then, is the anaphoric relation to be understood? And, more broadly, what is the logical form of such statements?⁵

In the present paper I will be arguing that when existential import is dropped from quantifiers, so that de dicto uses of (1) can be formalized as (2), a distinctively new puzzle appears. But before we turn to the new puzzle, it will be useful to consider some earlier types of reactions to Geach's original problem.

II. HISTORICAL BACKGROUND: RESPONSES TO GEACH'S PUZZLE

One possible response to Geach's puzzle is to argue against the linguistic intuitions that give rise to it. Not all ways of going about this are methodologically sound. It won't do to argue against the intuitions by pointing out that at present we lack a theory of intentional identity that would explain those intuitions. Linguistic intuitions don't always have to be defended, even by a crude theory, before they become admissible data for or against a semantical theory. In particular, requests for a general criterion of intentional identity needn't be met before one's intuition that a sentence like (1) is true in

circumstances like those in Examples 1 and 2 is admissible data. Nor will it do to argue against the recalcitrant intuitions by presenting cases for which it is hard to decide whether (1) is true. There are possible situations for which we cannot decide whether (8) below is true – how many grains of sand make a heap? – but (8) is actually true nonetheless.

(8) There exists at least one heap of sand.

Vagueness of various kinds is a fact of natural language we just have to learn to live with. That there are difficult cases thus does not even begin to show that *de dicto* (1) cannot be true in cases like the ones we've considered.

An argument of greater methodological appeal is offered against the recalcitrant intuitions in Montague [14], but I do not find the argument convincing. Montague suggests that if sentence (9) had a *de dicto* reading, then so would sentence (10); but the latter, he claims, has no such reading.

(9) John tries to find a unicorn and wishes to eat it.

(10) John wishes to find a unicorn and tries to eat it.

If this argument established that (9) lacked a *de dicto* reading in spite of our strong intuitions to the contrary, it would constitute significant evidence that intentional identity statements quite generally lack *de dicto* readings.⁶

The argument rests on three theoretical assumptions. The first can be expressed in the following way. Suppose (a) that phrases α and β are of the same syntactic category, (b) that under a given syntactic analysis, sentence S entails (does not entail) sentence Φ , and (c) that sentence S' results from S by replacing α for β . Then sentence S' , under a similar syntactic analysis, entails (does not entail) sentence Φ , if Φ contains neither α nor β .⁷ The second assumption is that 'tries to' and 'wishes to' are of the same syntactic category. The third theoretical assumption is that (10) has no *de dicto* reading; that is, on every admissible syntactic analysis, (10) entails 'A unicorn exists'. From these assumptions it follows that on every admissible syntactic analysis, (9) entails 'A unicorn exists'; that is, (9) has no *de dicto* reading.⁸

Evidence that the argument is unsound comes from certain variants of (10). Sentences (10a) and (10b) sound just as peculiar as (10) on any attempt to give them a *de dicto* reading (of both conjuncts).

- (10a) John wishes to find a unicorn and tries to eat the unicorn he finds.
- (10b) John wishes to find a unicorn and tries to eat the unicorn he wishes to find.

If on this basis we concluded that (10a) and (10b) lack *de dicto* readings, the first two assumptions of Montague's argument would then entail that (9a) and (9b) likewise lack *de dicto* readings.

- (9a) John tries to find a unicorn and wishes to eat the unicorn he finds.
- (9b) John tries to find a unicorn and wishes to eat the unicorn he tries [i.e., he is trying] to find.

But Montague's theory entails that (9a) and (9b) have *de dicto* readings, and most of us would agree with that. So something has gone wrong. Either (i) one of the first two assumptions of Montague's argument is mistaken, or (ii) the hypothesis that (10a) and (10b) lack *de dicto* readings is mistaken, and an alternative explanation of why they sound peculiar must be found. Surely, however, any reasonable such explanation could likewise be used to explain the strange sound of (10) itself, without the hypothesis that (10) lacks a *de dicto* reading – in which case the third assumption of Montague's argument would lose its plausibility.⁹

These considerations strongly suggest there is a problem with the argument, and I would locate it in the third assumption. The evidence that (10) lacks a *de dicto* syntactic analysis is too weak to support a conclusion as counterintuitive as the one that intentional identity statements lack *de dicto* readings. Certainly (10) sounds peculiar when you try to give it a *de dicto* reading, and it seems impossible to imagine a situation in which *de dicto* (10) would be true. Still, the assumption that (10) lacks a *de dicto* reading is not the only reasonable explanation of this data. Another perfectly reasonable hypothesis is that (10) *has* a *de dicto* analysis, but that on this analysis the

sentence is a psychologically necessary falsehood – a sentence false in every situation in which certain psychological laws hold. On this hypothesis, the bizarre sound of de dicto (10) is due only to the bizarreness of the circumstances under which it would be true. Some support for the hypothesis derives from the fact that there do seem to be exceedingly persuasive principles of commonsense psychology with which de dicto (10) would conflict. Intuitively, de dicto (11) and (12) are true; indeed it seems impossible to conceive of a case in which they would be false; and de dicto (11) and (12) intuitively entail the denial of de dicto (10).

- (11) If someone is trying to eat something, he seems to perceive it.
- (12) If someone seems to perceive something, he does not wish to find it.

Indeed, we might say this provides an intuitive *explanation* of why de dicto (10) must be false.¹⁰ So none of the attempts to undermine the intuitions that give rise to Geach's puzzle are successful.

A more attractive response to Geach's problem has been to construe the anaphoric relation in intentional identity statements quantificationally, but to drop existential import from the quantifiers used to formalize the de dicto uses. On this suggestion, (13) should be used to formalize de dicto (1), but it should not be used to formalize (14) (since the quantifier in (13) now carries no existential import).

- (1) Hob thinks a witch blighted Bob's mare, and Nob thinks she killed Cob's sow.
- (13) $(Ex)[\mathbf{B}_i(Wx \ \& \ Bx) \ \& \ \mathbf{B}_n(Cx)]$.
- (14) There exists an object such that Hob thinks it is a witch that blighted Bob's mare, and such that Nob thinks it killed Cob's sow.

(Sentence (14), and de re uses of (1), would be formalized using quantifiers carrying existential import, distinct from those used to formalize de dicto (1).)¹¹ Existential import could be dropped from the relevant quantifiers by one of (at least) two means.

- (A) The quantifiers could be interpreted substitutionally, while allowing non-denoting proper names and/or definite descriptions as substituends.
- (B) The quantifiers could be interpreted referentially, while allowing variables to range over both existent and non-existent objects. Non-existent objects could be either taken as primitive features of the semantic models or not. (In the latter case, the values of the variables could be partially defined individual concepts or world lines, in a possible worlds framework.)

Pendlebury [15] suggests that Geach's puzzle be solved by dropping existential import from quantifiers by method (A). Saarinen [16] suggests a solution in which existential import is dropped from the relevant quantifiers by method (B).¹²

It can be argued that method (A) offers no real advantage over method (B). Here I present only an informal version of the argument. On the most straightforward substitutional interpretation, an intentional identity statement such as (1) is true if and only if it has at least one true substitution instance, where a substitution instance of (1) is a sentence that results from it by replacing the quantifier phrase 'a witch' and its associated pronoun by a singular term drawn from a set of admissible substituends. The problem arises in attempting to specify a set of substituends that yields intuitively correct results.

Let's first assume that the set of admissible substituends includes natural language definite descriptions, and that these are given a Russellian analysis.¹³ It is easy to see that this will not work. Consider a situation such as the following one.

EXAMPLE 3. Joan decides to play a trick on Fred. She tells him that she has bought him a brand new Cadillac Coup de Ville, and that it is waiting for him in George's garage. In fact she has done no such thing, but Fred believes her. Purely by coincidence, Mabel decides to play the same trick on Charlie. She tells him a brand new Cadillac is waiting for him in George's garage. He believes her, but the garage is empty.

Intuitively, *de dicto* (15) could easily be false in the example, but *de dicto* (16) is true (on a Russellian analysis of the definite descriptions).

- (15) Fred believes a new car belongs to him, and Charlie believes it is a Cadillac.
- (16) Fred believes the new car in George's garage belongs to him, and Charlie believes the new car in George's garage is a Cadillac.

On the theory we are considering, however, (16) is a substitution instance of (15). So (15) could be false while one of its substitution instances is true. Obviously something is wrong.¹⁴

Suppose that we therefore restrict the set of admissible substituends to proper names. In order for this to work, for each true intentional identity statement the set of admissible substituends must contain a name that provides a true substitution instance when it replaces the relevant quantifier phrase and its associated pronoun. In simpler but less theoretically neutral language, there must be a name of each believed-in entity that is the focus of two or more people's beliefs, or that is the focus of different attitudes of the same person. Ideally, the admissible substituends would be unambiguous in the sense that, speaking again in less theoretically neutral language, no two believed-in entities have the same name. This would avoid the need for special provisions in the semantics for quantification to handle cases like the following.

EXAMPLE 4. Pat lives in Chicago, Ruth lives in Pittsburgh. Neither knows of the other's existence. In reality, neither Ruth nor Pat is married, but each believes herself to be married to someone. In both cases, there is no real person whom they believe to be their husband; the "husbands" are utter fantasy. Coincidentally, each woman believe her husband to be named "David Boswell".

Here, (17) could easily be false; but (18), on one of its uses, is true.

- (17) Pat believes someone lives in Chicago, and Ruth believes he lives in Pittsburgh.

- (18) Pat believes David Boswell lives in Chicago, and Ruth believes David Boswell lives in Pittsburgh.

If the set of admissible substituends contains names that are homonymous in the way that 'David Boswell' is in the example, and the semantical rule for quantification makes no provision for this, then a false intentional identity statement like (17) could have true substitution instances.¹⁵ Neither of these two requirements are met if we take ordinary proper names as the set of admissible substituends.

What is needed is a special category of names artificially introduced for the purpose of interpreting intentional identity statements. This is, in effect, Pendlebury's [15] suggestion. The idea is that whenever a *de dicto* intentional identity statement is true, a special (unambiguous) name is introduced into the set of admissible substituends to provide a suitable substitution instance. The problem with the suggestion is not only that it is *ad hoc* and inelegant, it is hard to see what real advantage it offers over the referential interpretation, which countenances non-existent objects of thought. All of the problems about the existence (or subsistence) and identity conditions for non-existent objects of thought are transposed into questions about the new artificially introduced names. Instead of asking under what conditions non-existent thought-objects "subsist", we now ask under what conditions the artificial names are introduced. Instead of asking under what conditions non-existent thought-objects are identical, we now ask under what conditions we should introduce two distinct artificial names rather than one. Nor does transposing these questions make them any easier to answer. Nothing is to be gained by gerrymandering things so that questions about metaphysics become questions about syntax, or more specifically, about lexicography.

The most frequently heard objection to method (B) for interpreting intentional identity statements is that it commits "us" to a dubious metaphysical theory: it commits us to an ontology containing non-existent objects, and perplexing questions can be asked about such an ontology. This objection rests on one or more methodological confusions. The claim that the (referential) quantificational interpretation of intentional identity ontologically commits us to non-existent objects depends on taking the variable as the linguistic mirror of ontic

commitment. The variable does reflect ontological commitment if you take its model-theoretic semantic values to represent EXISTING OBJECTS. But nothing forces us to assign this representational role to the set-theoretic values of the variables. (Those who would everywhere read ontic commitment into individual variables make a hasty generalization from the standard representational roles assigned to the domains of first order logic, in standard applications of that formal theory.)

Even if the quantificational interpretation did “commit us” to a dubious metaphysics, it is not at all clear that this would be the least reason for rejecting that interpretation. The question is *who* would be committed to the dubious metaphysics. If the quantificational interpretation forced the semanticist, *qua* semanticist, to countenance things he didn’t believe in, this would certainly be bad news for the theory. We don’t want the semantical metatheory to entail the truth of an implausible metaphysical doctrine. But a semanticist needn’t worry on this score. As long as he restricts himself to talking about algebraic structures used to explain certain inferential features of the language, he is on safe ground. Indeed, he can also say what various features of these algebraic structures represent, from the perspective of some metaphysical theory (perhaps not his own), without having to accept the ontological commitments of that metaphysical theory. The cultural anthropologist who explains why all the natives are prostrating themselves by saying “‘The Great Mountain God’ is near’ doesn’t thereby commit himself to the existence of the Great Mountain God, but he may thereby make us understand what the natives are doing, and why they’re doing it. So a semanticist, *qua* semanticist, is not in any danger of having to accept an obscure ontology because he accepts the (referential) quantificational interpretation. For those who *use* intentional identity statements, taking them to be literally true, it may be another story. If a speaker’s ontological commitments are reflected in his use of individual variables, then the quantificational interpretation has the consequence that those who use intentional identity statements “in earnest” are committed to an ontology of non-existent thought-objects. Things get tricky, however, when a formal semanticist is coincidentally a serious player of the intentional identity language game. In that case it may

be easy for him to confuse his ontological commitments as a serious player of the game with his commitments as a theoretician of the game. But it's essential to keep these distinct. A semanticist who, *qua* theoretician, criticizes a semantical theory on the grounds that it commits him, as a speaker of the object language, to an unpalatable metaphysics may well be in the position of a senator who argues against a congressional bill on the grounds that it will increase the cost of raw materials for his factory. So in my view the metaphysical arguments against the referential quantificational interpretation don't go very far.

To summarize our progress, we have examined Geach's original puzzle and a few unsuccessful types of attempts to dismiss it. Our discussion suggests that of the strategies we've considered for interpreting intentional identity, the most promising is to offer a quantificational analysis in which existential import is dropped from the relevant quantifiers. Although the substitutional interpretation of these quantifiers holds no special appeal in the present context, the usual "metaphysical" arguments against interpreting them referentially are methodologically suspect. I'll be arguing in the next section, however, that a new puzzle demonstrates the impossibility of a quantificational interpretation of natural language intentional identity.

III. THE NEW PUZZLE

I wish to discuss a more strictly logical problem besetting the quantificational interpretation of intentional identity. The difficulty arises for both the referential and substitutional interpretations of the relevant quantifiers. The problem is that formalizing intentional identity statements quantificationally fails to preserve the right consequence patterns for these sentences, on any otherwise well-behaved semantics. More precisely, quantificational interpretation of intentional identity has the result that inferences not valid in English have formal translations that are valid on any semantics in which conjunction is symmetric (when not in the scope of propositional attitude operators); that is, on any semantics in which the following inference is valid when sentences Φ and Ψ do not occur in the scope of any such operators.

... Φ & Ψ ...

... Ψ & Φ ...

I don't require that the inference above hold when Φ and Ψ occur in propositional attitude contexts, simply in order to emphasize that the problem arises even when the semantics fails to preserve the validity of these inferences within belief contexts. The problem for the quantificational interpretation is *not* that people do not always believe all the logical consequences of what they believe.

There are a number of inferences that are an embarrassment to the quantificational interpretation. The only one I will consider here concerns what might reasonably be called the asymmetry of intentional identity. Consider for instance the inference from (19) to (20) below.

- (19) Detective A believes someone murdered Smith, and
Detective B believes he murdered Jones.
- (20) Detective B believes someone murdered Jones, and
Detective A believes he murdered Smith.

First I wish to argue that the inference is intuitively invalid; later we will consider its quantificational interpretation.

EXAMPLE 5. Two detectives, A and B, recently investigated an apparent murder on Chicago's south side. Smith's body had a bullet hole in it, and Detectives A and B inferred that Smith was murdered by a single person. They discussed the case at length, but neither has anyone in mind as a suspect.

Yesterday the two detectives were investigating another apparent murder, this time on Chicago's north side. Jones' body had a bullet hole in it, and Detectives A and B inferred that Jones was murdered by a single person. Detective B thinks that the man who murdered Smith is the same person as the man who murdered Jones. Detective A disagrees. He thinks Smith and Jones were murdered by two different people, though he has no one in mind as a suspect for either case. The two detectives argue heatedly about whether the man who murdered Smith is the same person as the man who murdered Jones. (For purposes of later discussion, we assume that Detective A believes that Smith's and Jones' murderers are both still in Chicago.)

There can be no doubt that on its most natural use, (19) is true in the described situation. Probably one already has some strong doubt about whether the most natural reading of (20) is true. It is easy to fill out the example in such a way that this doubt reaches certainty.

EXAMPLE 6. This is just like Example 5, except that in fact neither Smith nor Jones was really murdered. Both died of heart attacks. In each case, a stray bullet from an unrelated incident struck the corpse in the chest.

This last example shows that the inference is invalid even when the two sentences are given their *de dicto* reading; that is, even when the speaker is not committed to the actual existence of a common focus of the two detectives' beliefs.¹⁶

It is easy to see that the quantificational formalization of the inference will be valid on any semantics in which conjunction is symmetric at least outside the scope of propositional attitude operators. For interpreted quantificationally, the inference will be formalized as follows.

$$(21) \quad (Ex)[\mathbf{B}_a(Sx) \ \& \ \mathbf{B}_b(Jx)]$$

$$(22) \quad (Ex)[\mathbf{B}_b(Jx) \ \& \ \mathbf{B}_a(Sx)].$$

(22) follows from (21) by one application of the principle of the symmetry of conjunction.

Although it may be possible to develop formal semantical theories in which conjunction is not symmetric, the unmanagability of the resulting logic would certainly speak against it. Surely the source of the problem lies not in the treatment of conjunction in standard intensional logics, but in the attempt to interpret intentional identity statements quantificationally. The problem arises, in other words, from the assumption that the pronouns in intentional identity statements are quantificational pronouns.

It is interesting that the laziness of interpretation of intentional identity provides an elegant explanation of the breakdown of the English inference from (19) to (20) in Examples 5 and 6. On the laziness analysis, (19) and (20) could be taken to abbreviate (23) and

(24), respectively, where as usual the definite descriptions take narrowest possible scope.

- (23) Detective A believes someone murdered Smith, and Detective B believes the person who murdered Smith murdered Jones.
- (24) Detective B believes someone murdered Jones, and Detective A believes the person who murdered Jones murdered Smith.

Thus understood, (23) and (24) are conjunctions of the de dicto belief statements. Obviously, de dicto (23) does not entail de dicto (24). Moreover, standard intensional logic provides an explanation of the intuition that the inference from de dicto (23) to de dicto (24) is invalid. So if the pronouns in (19) and (20) are pronouns of laziness, we then have a facile explanation of the failure of the inference from the former to the latter: the recalcitrant uses of (19) and (20) are mere abbreviations of de dicto (23) and (24), and the former quite comprehensibly does not entail the latter.

One will recall from Section I that in at least some cases, the pronoun in an intentional identity statement cannot be interpreted as a pronoun of laziness. This might well make one suspicious of the laziness explanation of the breakdown of the inference from (19) to (20). But it would be unfair to rule that explanation out of hand. The laziness explanation is after all perfectly compatible with each of Examples 5 and 6. In each of the two examples, a sentence (19) could plausibly abbreviate, namely (23), is true; and a sentence (20) could plausibly abbreviate, namely (24), is false.

The success of the laziness interpretation in explaining why the inference breaks down in Examples 5 and 6, and the failure of the laziness interpretation for the cases considered in Section I, might lead one to propose that the pronouns in de dicto intentional identity statements are ambiguous between a quantificational use and a lazy use. Such a suggestion would most definitely *not* be ad hoc. There is independent evidence that certain pronouns are ambiguous in precisely this way. For instance, Karttunen's [11] sentence (25) below is ambiguous between a reading on which it is talking about two

paychecks, and one on which it is talking about a single paycheck. On one reasonable hypothesis, the pronoun in the first use is a pronoun of laziness; the pronoun in the second use is a quantificational pronoun.

- (25) The man who gave his paycheck to his wife was wiser than the man who gave it to his mistress.

One wonders, then, whether there are cases for which the inference breaks down, but which also preclude the laziness interpretation of both sentences. In fact there are such examples. Unfortunately they are a bit more complicated than the examples used to show that the inference breaks down; but this is to be expected since such examples must contain not only those features which lead to the breakdown of the inference, but also those features which block the laziness interpretation. We need to consider only a slightly more complicated version of our inference.

- (26) Detective A thinks someone murdered the mayor, and Detective B thinks he murdered the commissioner.
- (27) Detective B thinks someone murdered the commissioner, and Detective A thinks he murdered the mayor.

Example 7 is an “anti-laziness” counterexample to the inference from (26) to (27).

EXAMPLE 7. *Monday:* Smith (the mayor) and Jones (the commissioner) have been shot, at opposite ends of Chicago. Detectives A and B are investigating both cases, but neither knows that Smith is the mayor or that Jones is the commissioner. Smith and Jones, though hospitalized, are (and are known by both detectives to be) still alive. A and B have discussed the two cases at length, and though they think someone shot Smith and that someone shot Jones, both believe the two cases are entirely unconnected. At this time, neither has anyone in mind as a suspect.

Tuesday: Both Smith and Jones have died of their gunshot wounds. Detective A knows Smith died, and thus now believes that the person

who shot Smith murdered him, but doesn't know Jones is dead. Likewise, B knows Jones died, and thus now believes that the person who shot Jones murdered him, but doesn't know Smith is dead. Detective A now knows that Smith was the mayor, but not that Jones was the commissioner. Similarly, B now knows that Jones was the commissioner, but not that Smith was the mayor. After reflecting on certain similarities between the two cases, Detective B infers that the man who shot Smith is the same person as the man who shot Jones. He communicates this to A, saying, "The man who shot Smith is the man who shot Jones." A disagrees, but B persists in his opinion.

On Tuesday, (26) on its most natural reading is true. But on Tuesday, (27) on its most natural reading is false. The details of the example obviously preclude a laziness analysis of either sentence: B believes neither that Smith was the mayor nor that he was murdered; A believes neither that Jones was the commissioner nor that he was murdered. So the laziness interpretation cannot be used to explain the breakdown of the inference from (26) to (27).

IV. SIGNIFICANCE OF THE NEW PUZZLE

If the arguments just given are sound, neither the laziness *nor* the quantificational interpretations of intentional identity are acceptable. The anaphoric relation in (at least some) intentional identity statements must be construed in some alternative fashion.

From one perspective, this result might not seem terribly interesting. Philosophers and linguists have recently argued that other anaphoric phenomena are beyond the scope of the quantificational and laziness approaches; it's not surprising, then, that an approach that fails elsewhere should likewise fail for intentional identity.¹⁷ In the present section I will be arguing that the failure of the symmetry inference for intentional identity statements raises problems for the interpretation of propositional attitudes beyond those specifically concerned with anaphora. I will be arguing that theories able to explain the breakdown of similar symmetry inferences for purely extensional discourse cannot by themselves explain the breakdown of the symmetry inference in the case of intentional identity. For the sake of

simplicity, the argument is here couched in a framework for anaphora simpler than the one I would ultimately prefer, but the argument generalizes to more sophisticated frameworks in which the basic ideas are preserved.

It has recently been argued that an anaphoric pronoun can have as its semantic value the “speakers referent” of its antecedent (Donnellan [2], Kripke [12], p. 21). If this is correct, it would perhaps *seem* as if this fact alone would yield an acceptable explanation of the breakdown of the inference. After all (one might argue) inferences such as the ones we have been considering can break down even in purely extensional contexts. Suppose the speaker of (28) is speaker-referring to Jackie in uttering the pronominal antecedent, and the speaker of (29) is speaker-referring to Charlotte in uttering the pronominal antecedent.

(28) Someone is in the park, and she is eating an apple.

(29) Someone is eating an apple, and she is in the park.

If in certain uses of these sentences, the pronoun can acquire as its *semantic* referent the speaker’s referent of its antecedent, then the semantic meaning of (28) could be true while the semantic meaning of (29) is false (e.g., if Jackie is in the park eating an apple, but Charlotte is neither in the park nor eating an apple).

One might then attempt to explain the breakdown of the inference from (19) to (20) along the same lines. It can be seen, however, that the breakdown of the inference from (19) to (20) generates problems not present in the purely extensional cases like (28) and (29). This becomes evident once one tries to work out the details of an explanation of the breakdown of the inference for intentional identity as we did for the extensional case. The difficulties become apparent even in the absence of a detailed and precise theory.

Suppose we say that the things people believe in are *thought-objects*. Let’s allow for the present that a single thought-object may exist in more than one person’s beliefs (later we will see what happens when this assumption is abandoned). Thus we say that the thought-object *Santa Claus* exists in many different people’s beliefs: he is, we might say, *intersubjectively extended*. (I use star-quotes to form names

of thought-objects.) We might also allow that a thought-object can exist both in people's beliefs and in the real world, when the thing they believe in exists, but this detail needn't concern us here. I assume that the relation of intersubjective identity is an equivalence relation, at least for the present. In this framework, an intentional identity statement will be true if there is an intersubjectively extended thought-object of the right sort: one that exists in the beliefs of both people, and which has the right properties in those people's beliefs.

How, in the context of such a framework, can we account for the breakdown of the inference from (19) to (20)? Clearly, more than one thought-object will be required if we are going to explain the breakdown of the inference as we did for the one from (28) to (29). Just as we needed to appeal to two distinct objects, Charlotte and Jackie, to serve as the semantic referents of the pronouns to explain how (28) could be true while (29) is false, we need to appeal to two distinct thought-objects to serve as the semantic referents of the pronouns in (19) and (20). This is where the real problems begin.

Suppose we say that in Examples 5–7, there are at most two intersubjectively extended thought-objects which for convenience we can refer to as *the man who shot Smith* and *the man who shot Jones*.¹⁸ Are these the same thought-object or not? There are serious problems whichever way we answer. If we say yes, we can account for the truth of (19) in the examples; but we won't be able to account for the falsity of (20) since there won't be a second thought-object to serve as the semantic referent of the pronoun in (20). If on the other hand we say that these are distinct thought-objects, we will not be able to account for the false use (30) below has in Examples 5 and 6.¹⁹

- (30) Detective B believes someone murdered Smith *and* Jones, and Detective A believes *he* is still in Chicago.

For in uttering (30), the speaker could then be speaker-referring only to one of the intersubjectively extended thought-objects, *the man who shot Smith* (who in B's beliefs did murder Smith and Jones, and who in A's beliefs is still in Chicago), or *the man who shot Jones* (who in B's beliefs did murder Smith and Jones, and who in A's beliefs is still in Chicago).

So far we have assumed that thought-objects can be intersubjectively extended. *The man who shot Smith* can exist in both A's and B's beliefs. One might think that this is the source of our difficulty. Let's assume, then, that thought-objects can exist only in a single person's beliefs, and that thought-objects in various people's beliefs can be *counterparts* of one another. Thus Detective A's *the man who shot Smith* is not identical to, but is a counterpart of, Detective B's *the man who shot Smith*. We will assume that the counterpart relation is reflexive but neither symmetric nor transitive. The pronominal antecedent in an intentional identity statement can speaker-refer to such a "person-bound" thought-object. In such cases the semantic value of the pronoun is then the counterpart (if there is one) in the second person's beliefs of the thought-object referred to by the speaker in uttering the pronoun's antecedent (otherwise the pronoun carries no semantic value).

A theory of this sort solves the problems considered so far, but is subject to other difficulties. A natural interpretation of Examples 5–7, in the current framework, is that Detective A's beliefs contain two thought-objects, *A's the man who shot Smith*, and *A's the man who shot Jones*; Detective B's beliefs contain only the thought-object *B's the man who shot Smith and Jones*. We can abbreviate the names of these three thought-objects, respectively, as [A-sS], [A-sJ], and [B-sSJ]. Let *R* be the counterpart relation. Then it is reasonable to assume the following.

[A-sS] bears *R* to [B-sSJ] but not vice versa.

[A-sJ] bears *R* to [B-sSJ] but not vice versa.

[A-sS] does not bear *R* to [A-sJ], nor vice versa.

In that case, we can explain the true use of (19): 'someone' speaker-refers to [A-sS], 'he' semantically refers to [B-sSJ]. Likewise, we can also explain the false uses of (20) and (30): in both cases, 'someone' speaker-refers to [B-sSJ], but 'he' lacks a semantic referent since [B-sSJ] bears *R* to no thought-object in A's beliefs. But we will have held down the demon in one place only to have him pop up in another. For now we will be at a loss to explain the *true* use (31) has in Examples 5 and 6.

- (31) Detective B thinks someone (Smith's murderer) murdered Smith, and Detective A thinks he is still in Chicago.

Altering the counterpart relations described above so as to allow for the true use of (31) (by replacing 'but not' to 'and' in the first two *R*-statements above) would result in our being unable to account for the false use of (20). Introducing a non-symmetric counterpart relation, in other words, doesn't really help. So we are still without a solution to the new puzzle. It follows that we are still without an *adequate* solution to the puzzle Geach proposed in 1967.

In closing I wish to sketch the direction in which I think an adequate solution lies. I am not going to present a full-blown theory (let alone a formal one), but only make a few suggestive remarks. It seems to me that the source of the problems we have been discussing is that we have been operating with an impoverished interpretation of the examples. Let's return to the framework of intersubjectively extended thought-objects (this only makes for briefer exposition; it's not an essential part of the suggestion). We ran into problems when we interpreted the examples as containing only the two thought-objects, *the man who shot Smith* and *the man who shot Jones*. I suggest that the examples are best construed as containing three intersubjective thought-objects: *the man who shot Smith*, *the man who shot Jones*, *the man who shot Smith and Jones*. The first two thought-objects exist in both A's and B's beliefs, the third only in B's beliefs. (Of course the first two thought-objects bear a very special relation to the third, as these exist in B's beliefs. In a technical sense yet to be explicated, they are "part" of it.)²⁰

This analysis of the examples offers a reasonable hope that we can account for the data we've been discussing. In true uses of (19), the speaker refers to *the man who shot Smith* (in A's beliefs, this object did murder Smith, and in B's beliefs it murdered Jones). In false uses of (20), the speaker refers to *the man who shot Smith and Jones*, or to *the man who shot Jones* (the former object does not exist in A's beliefs, and the latter did not murder Smith in A's beliefs). We can account for the false use of (30) by saying that the speaker refers to *the man who shot Smith and Jones* (which doesn't

exist in A's beliefs). Finally, in the true use of (31), the speaker refers to *the man who shot Smith* (in B's beliefs, this object did murder Smith, and in A's beliefs, it is still in Chicago).

Of course this does no more than gesture at the direction I think an adequate solution lies, and leaves us with a number of unresolved questions. How should we understand thought-objects? What is it for them to be intersubjectively extended (or for them to be counterparts, in the alternative version of the framework)? What is the relation between thought-objects like *the man who shot Smith* and *the man who shot Smith and Jones* as these exist in B's beliefs? What would the formal theory of the type of solution suggested here look like? Interesting and urgent as these questions may be, they are well beyond the scope of the present paper.²¹ My intention here has only been to present the puzzle, and state where I think an adequate solution is most likely to be found.

NOTES

* I wish to thank Nuel Belnap, Charles Chastain, Dorothy Grover, Anil Gupta, and Gerald Massey for their kindness in reading and commenting on an earlier draft of this paper. Special thanks are due to Joseph L. Camp, who first introduced me to the problem of intentional identity, and who profoundly influenced my thinking on this and related issues.

¹ Geach's original sentence was 'Hob thinks a witch has blighted Bob's mare, and Nob wonders whether she (the same witch) killed Cob's sow.' In the interest of simplicity, I consider a sentence having the same psychological verb in both clauses. The problem sentences typically contain two separate attitudinal contexts, one of which contains a pronoun whose antecedent lies in the other.

² Sometimes alternative criteria are used to draw the (or rather, a) *de dicto/de re* distinction. I do not mean to suggest that the way the distinction is drawn here captures every distinction that has been called the *de dicto/de re* distinction, nor do I mean to suggest that the criterion employed here is in any sense more correct than certain other criteria. Notice that as I have drawn the distinction here, it is not a theoretical one. I have merely drawn a distinction between uses of a sentence by means of intuitions about inference patterns, and attached labels to those two uses. Some uses of (1) entail 'A witch exists' or at least 'There exists an object about which Hob and Nob both have beliefs'; others do not. However these two uses are to be formally interpreted, I call the former uses "de re", the latter, "de dicto".

³ I say *largely* a syntactic hypothesis because whether an expression is a quantifier (or quantifier phrase) depends in part on the kind of semantic treatment it gets, and not entirely on its syntactic behavior.

⁴ See Geach [5] and [6]. Strictly speaking, Geach's original definition of a pronoun of

laziness is narrower than this. In Geach [5] such a pronoun must go proxy for an expression syntactically identical to its antecedent, but in [6] he broadens the definition to include pronouns that go proxy for expressions constructible from words in the environment of the antecedent.

⁵ This characterization of the puzzle neglects some important aspects of its significance. One reason Geach's puzzle is philosophically interesting is that it shows that standard intensional logic, in which existential quantifiers carry existential import, is of limited explanatory utility in an important respect. It has been thought that standard intensional logic provides a theoretical framework within which the *de dicto/de re* distinction, drawn in terms of intuitions about existential commitments as we have done here, could be explained as a distinction of scope (of quantifiers and intensional operators). Geach's puzzle shows that this explanation is inadequate for at least *some* instances of the intuitively drawn *de dicto/de re* distinction. This seriously jeopardizes the plausibility of using the scope distinctions of standard intensional logic to account for the distinction in the more familiar cases.

⁶ After giving this argument, Montague goes on to say that it is possible to interpret the pronoun in (9) as a pronoun of laziness and thereby obtain a *de dicto* analysis of the sentence. Since he has already argued against the intuition that (9) has a *de dicto* reading, I assume he is here merely stating how those who are unpersuaded by the argument could accommodate this intuition in a way consistent with his theory.

⁷ This assumption can easily be made more precise in Montague's framework by utilizing the notion of an analysis tree to clarify what is meant by a "similar syntactic analysis". Such a detailed statement of the principle is not required for our present purposes, however.

⁸ In interpreting Montague's argument I have assumed that by 'reading' he means 'syntactic analysis'. An alternative interpretation would assume that by 'reading' Montague means 'semantical analysis', but this would affect neither the structure of the argument nor the nature of my criticisms of it.

⁹ One will notice that in effect my objection argues from uses of (9) and (10) in which the pronoun is a pronoun of laziness. I once thought it possible to offer a similar objection, based on *de re* uses of (9) and (10). For at first it appears that (10) is as peculiar on its *de re* reading as on its *de dicto* reading (you can't try to eat something you still wish to find). If on this basis we concluded that (10) lacks a *de re* reading, Montague's first two principles would then commit us to saying that (9) likewise lacks a *de re* reading — a conclusion few would accept. I now have doubts about the possibility of such an objection based on *de re* uses. For in fact *de re* uses of (10) are not as peculiar as its *de dicto* uses: it is possible to conceive of circumstances in which *de re* uses would be true. Suppose that a real unicorn, Tom, exists, and that John wants to find Tom and eat him. John finds Tom and then tries to eat him, but doesn't realize it's Tom he's found: he still wishes to find Tom. Here *de re* (10) would be true. It seems impossible to conceive of an analogous instance involving fictional or merely imagined unicorns, in which *de dicto* (10) would be true.

¹⁰ Of course sentences (11) and (12) are themselves intentional identity statements, and Montague's argument purports to show that contrary to our intuitions these have no *de dicto* reading. The situation, then, is this. It is possible to accept Montague's argument at the cost of intuitions concerning *de dicto* (9), (11), (12), and similar sentences (and, arguably, at the cost of intuitions about *de dicto* (9a) and similar sentences). Alternatively this range of intuitions can be preserved at the cost of one of the

premisses of Montague's argument. Both alternatives provide an explanation of the peculiar sound of *de dicto* (10), but the latter one preserves a far wider range of linguistic intuitions.

¹¹ Alternatively, the same set of quantifiers could be used to formalize *de dicto* (1) and (14), and a pragmatic component could be used to determine whether a given existential quantifier carries existential import in a given context.

¹² Drawing on the same work of Hintikka's that led Saarinen to his solution, I arrived at essentially the same theory in 1978. See Hintikka [8], [9], and [10]. As will become evident in the next section, I now believe such a theory is mistaken.

¹³ Dennett [1] argues for an interpretation that amounts to this one (though he allows that in some uses the pronouns may be functioning as pronouns of laziness). Geach considered such an interpretation in [7] and rejected it, though his argument is different from the one presented below.

¹⁴ It doesn't really help to try to obtain a false reading of (16) by interpreting the second definite description in such substitution instances in a non-Russellian way as an anaphor functioning like a pronoun whose antecedent is the first definite description. Though (16) does have such a use, it only raises all over again the problem of how to understand anaphoric relations between separate attitudinal contexts.

¹⁵ Obviously the two requirements stated in this paragraph have analogs even in more mundane applications of substitutional quantification, in which the language is used to talk about only existing objects. You must have a name of every real object (or else a "renaming" device in the semantical clause for quantification that gives the same effect) And you must ban homonymous names from the class of substituends (or include a pragmatic component in the rule for quantification, in order to handle them properly).

¹⁶ I think there is also a use of (19) on which it is false in these examples, and a use of (20) on which it is true. If there are such uses, then there are corresponding uses of the inference on which it is an admissible one to make. Still, on their most natural uses in the examples, (19) is true and (20) is false; and on its most natural use in the examples, the inference is a bad one to make. At present it is this latter intuition I wish to account for.

¹⁷ Evans [4], pp. 492–520, contains a nice collection of such arguments. See also Donnellan [2], §4.

¹⁸ We do not assume that *the *F** is *F* in every person's beliefs in which it exists (otherwise we may run into trouble with cases similar to Example 7). Rather, '*the *F**' is a convenient tag for the intersubjectively extended thought-object which, in at least one person's beliefs, is *F*.

¹⁹ I think sentence (30) has both true and false uses in Examples 5 and 6. One context in which the false use is more likely is one in which the speakers in the context are discussing disagreements about the number of murderers involved, and their present whereabouts.

²⁰ In a wide range of cases, the notion of explanatory role seems useful in explicating the relevant part–whole relation: Thought-object *X* is part of thought-object *Y*, in the relevant sense, iff every explanatory role played by *X* is played by *Y*. This does seem appropriate for the examples considered here, but it is merely a suggestive remark and is not intended to cover all the sorts of cases that can arise, such as intentional identity statements about mathematics ('John thinks the solution to the problem posed by the mayor is even, but Mary thinks it's odd').

²¹ I have given some attention to these questions in [3].

REFERENCES

1. Dennett, Daniel, 'Geach on Intentional Identity', *Journal of Philosophy* **65** (1968), 335–341.
2. Donnellan, Keith, 'Speaker Reference, Descriptions, and Anaphora', in P. Cole (ed.), *Syntax and Semantics 9*. Reprinted in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, University of Minnesota Press, Minneapolis (1979) pp. 28–44.
3. Edelberg, Walter, *Intentional Identity*, Dissertation, University of Pittsburgh (1984).
4. Evans, Gareth, 'Pronouns, Quantifiers, and Relative Clauses', *Canadian Journal of Philosophy* **7** (1977), 467–797.
5. Geach, Peter, *Reference and Generality*, Cornell University Press, Ithaca (1962).
6. Peter Geach, 'Referring Expressions Again', *Analysis* **24** (1963–64). Reprinted in P. Geach, *Logic Matters*, Blackwell, Oxford (1972), pp. 97–102.
7. Peter Geach, 'Intentional Identity', *Journal of Philosophy* **74** (1967). Reprinted in P. Geach, *Logic Matters*, Blackwell, Oxford (1972), pp. 146–153.
8. Hintikka, Jaakko, 'On the Logic of Perception', in N. Care and R. Grimm (eds.), *Perception and Personal Identity*, Cleveland (1969), pp. 140–173.
9. Hintikka, Jaakko, 'Objects of Knowledge and Belief: Acquaintances and Public Figures I', *Journal of Philosophy* **67** (1970) 869–884.
10. Hintikka, Jaakko, 'Knowledge by Acquaintance – Individuation by Acquaintance', in D. F. Pears (ed.), *Bertrand Russell: A Collection of Critical Essays*, Garden City, NJ (1972), pp. 52–79. Reprinted in J. Hintikka, *Knowledge and the Known*, D. Reidel, Dordrecht (1972), pp. 212–233.
11. Karttunen, Lauri, 'Pronouns and Variables', in R. I. Binnick *et al.* (eds.), *Papers from the Fifth Regional Meeting of the Chicago Linguistic Society*, Dept. of Linguistics, University of Chicago, Chicago (1969).
12. Kripke, Saul, 'Speaker's Reference and Semantic Reference', in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language*, University of Minnesota Press, Minneapolis (1979), pp. 6–27.
13. Kripke, Saul, 'A Puzzle about Belief', in A. Margalit (ed.), *Meaning and Use*, D. Reidel, Dordrecht (1979).
14. Montague, Richard, 'The Proper Treatment of Quantification in Ordinary English', in J. Hintikka, J. Moravcsik, and P. Suppes (eds.), *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, D. Reidel, Dordrecht (1973). Reprinted in R. Montague, *Formal Philosophy*, edited by R. H. Thomason, Yale University Press, New Haven (1974), pp. 247–270.
15. Pendlebury, Michael, 'Hob, Nob, and Hecate: The Problem of Quantifying Out', *Australasian Journal of Philosophy* **60** (1982), 346–355.
16. Saarinen, Esa, 'Intentional Identity Interpreted: A Case Study of the Relations Among Quantifiers, Pronouns, and Propositional Attitudes', in E. Saarinen (ed.), *Game-Theoretic Semantics*, D. Reidel, Dordrecht (1978).

Philosophy Department,
University of Illinois at Chicago,
Box 4348,
Chicago, IL 60680,
U.S.A.