Mutt and Jeff agree on what sentences Odile accepts. They agree about her dispositions to behavior. They agree on just about everything which seems relevant to the question, does Odile believe that Twain is dead?

They don't agree on the answer. When Mutt was asked, it was because someone wanted to know whether Odile would list Twain under dead Americans. Mutt knew she accepted 'Twain is dead' and thus said 'yes'. Jeff was asked by someone who couldn't understand why Odile, who's pointing to Twain's picture, wants to meet him. Doesn't she realize that Twain is dead? Jeff knew she rejected 'he's dead'. He answered that, no, Odile didn't believe that Twain was dead.

What are we to make of this? This paper investigates a way of saying that they're both right. Not because

(1) Odile believes that Twain is dead

is syntactically ambiguous. Not because there is a semantic ambiguity in the sentence. At least, (1) is not semantically ambiguous in the way that, say,

(2) Odile dropped Marie Bernard

is.

I propose that 'believes' and other verbs of propositional attitude are indexical. The truth of (1) varies across Mutt's and Jeff's contexts. There is not a change in reference in expressions other than 'believes', nor any change in Odile. And in some important sense 'believes' remains constant in meaning. If we accept all this then we will say that 'believes' is an indexical.

© 1989 by Mark Richard.
What varies across contexts that is relevant to the interpretation of 'believes'? I think that it's what counts as an acceptable translation of the sentences (in a very broad sense of 'sentence') Odile accepts; what varies is what functions are acceptable translation or correlation functions. (1) is true in a context iff its content sentence is an acceptable translation of some sentence Odile accepts. In Mutt's context, we may suppose, there are no substantive restrictions on translation at work: If Odile accepts "a is dead" for some name a of Twain, that makes (1) true, since 'Twain is dead' here translates any such sentence. In Jeff's context, something more stringent is required. Perhaps it's required that Odile accept "I am pointing at a" before 'Twain' can translate a.

I sketch some of the details of this view in section 1. The claim that attitude ascription involves translation will remind cognoscenti of Church's objections to Carnap's and other translational accounts of such sentences. In section II, I turn to consider how the view fares in the face of some Church-style objections. Section III discusses Kripke's Pierre-London and Peter-Paderewski cases. I conclude with a discussion of some broadly logical issues.

I'm going to make a somewhat controversial assumption: that belief and the other attitudes are had in virtue of relations to sentence-like entities, whose constituents determine (relative to a context) Russellian referents. That is, when I have a belief, I'm related to a "sentence" which typically contains things which function like natural language proper names and demonstratives (in determining an individual), predicates (in determining properties or relations), as well as constituents which have the semantic roles of natural language connectives, quantifiers, and so on.

In fact, for simplicity, I am going to assume that attitudes like belief are realized by relations to natural language sentences. The account I'll give could get by with a considerably weaker assumption than this—for one thing, the mediators of belief needn't be sentences of a natural language. But I'll leave a discussion and defense of this assumption for another occasion.¹

I thus assume a picture of the attitudes much like that assumed by many contemporary Russellians, who see belief as a relation to a Russellian entity, a relation mediated by relations to sentences or sentence-like entities.² I think this picture is a good first approximation to the truth about belief. But I am now unsatisfied with the account of belief ascription which the Russellian offers, on which 'Odile believes that Twain is dead' and 'Odile believes that Clemens is dead' must agree in truth value, since their 'that'-clauses name the same Russellian proposition.

What shall we offer in its place? Consider the example of Hammurabi, who reputedly believed that Hesperus was Hesperus, but not that Hesperus was Phosphorus. If we allow ourselves to talk about belief blackboards and
such, we can describe what the relevant facts are: There are two possible mediators of Hammurabi's belief—pretend that they are

(1) $H = H$

and

(2) $H = P$.

Here, $'H'$ is doing duty for the Babylonian word canonically translated 'Hesperus'; $'P'$, for that canonically translated 'Phosphorus', and $'=' $ goes proxy for a Babylonian identity predicate. Hammurabi, presumably, had (1) written on his blackboard, but not (2). It is this fact, or one like it, that someone who utters

(3) Hammurabi believes that Hesperus is Hesperus; but

(4) Hammurabi doesn't believe that Hesperus is Phosphorus

is trying to get across. The question is, how do (3) and (4) get this across?

In some way or other, in uttering (3) and (4) the speaker uses 'Hesperus' to represent $'H'$, 'Phosphorus' to represent $'P'$ and 'is' to represent $'= '$. And thus, the 'that'-clause 'that Hesperus is Hesperus' represents (1), while the 'that'-clause 'that Hesperus is Phosphorus' represents (2). (3) is true just in case what its t-clause represents is on Hammurabi's belief blackboard; likewise for (4). Since what (3)'s 'that'-clause represents is on the board, while what (4)'s represents isn't, (3) is true while (4) is not.

It will be said, perhaps, that this is correct, but that it leaves unanswered the most important question: How does 'Hesperus' come to represent $'H'$?

A popular answer is that 'Hesperus' and $'H'$ have the same or similar nonreferential, cognitive content: they have the same or similar Fregean senses or conceptual roles (for the speaker and Hammurabi, respectively).

I've argued elsewhere that this sort of answer is untenable. It is untenable because the sense (conceptual role, etc.) of a sentence will, on any reasonable account, vary intersubjectively. Here's a hint of the sort of problems which arise because of intersubjective variation. Observe that it's obvious that, if $x$ can use sentence $S$ to express a belief, and I can use $S$ without altering the references of its expressions, then I can use $S$ to ascribe to $x$ the belief she expresses with $S$. If, for example, Odile can express a belief about Twain with 'Twain is dead', then I could echo her words, and use 'Odile believes that Twain is dead' to ascribe a belief to her.

Now this is so, even if Odile and I associate wildly divergent senses with the sentence 'Twain is dead'. Odile needn't grasp my sense for the sentence, for my ascription to be true. And what's more, when I echo someone's words to ascribe belief or assertion to them, I never seem to worry about identity or similarity of sense between their words and mine.

These facts refute Frege's account of attitude ascription. He held that in 'Odile thinks that Twain is dead', the 'that'-clause names a sense, my sense for
‘Twain is dead’; the sentence as a whole says that Odile believes the sense named. But as we saw, the ascription may be true even though Odile doesn’t believe my sense for ‘Twain is dead’.

We were wondering how ‘Hesperus’ in ‘Ham believes that Hesperus is Hesperus’ comes to represent one or another of the expressions in Ham’s belief mediators. If we reject Fregean and allied accounts of attitude ascription, we will agree that the right answer to the question is not: ‘Hesperus’ has for us the same sense, or cognitive role, as ‘H’ for Hammurabi. Is, perhaps, the fact that one of the expressions conventionally translates the other relevant here?

Well, of course it’s relevant. But it is also something which is an accidental feature of the example. Recall the case of Mutt, Jeff, and Odile. Jeff is asked if Odile believes that Twain is dead. His inquisitor wonders whether Odile, pointing at Twain’s picture, realizes that Twain’s dead. Jeff knows that she does not accept ‘that one is dead’; he says ‘Odile doesn’t believe that Twain is dead’. Jeff’s use of ‘Twain’ represents Odile’s use of ‘he’ or ‘that one’. So it can represent ‘t’ without ‘t’ being a conventional translation of ‘t’, or vice versa.

In this case there doesn’t seem to be any interesting connection—beyond sameness of Russellian referent—between Jeff’s use of ‘Twain’ and the part of Odile’s belief mediator it represents. Beyond the identity of Russellian interpretation of ‘Twain’ and ‘that one’, there doesn’t seem to be anything about the content of ‘Twain’—in an intuitive sense of ‘content’—which makes it an apt representative of Odile’s use of ‘that one’. The two needn’t have the same cognitive content; neither need the one be a translation, in some meaty sense of ‘translation’, of the other, etc.

The idea that there is something intrinsic to the content of Jeff’s sentence, which would make it represent one but not another of the mediators of Odile’s belief, is a mistake. ‘Twain’ can in principle represent any name, demonstrative or indexical, which Odile uses to refer to Twain. Of course, given contextual factors, there may be something which makes ‘Twain’ a more apt representative of ‘that one’ than other expressions the speaker might have used. Given the context of Jeff’s remark, it’s clear to all—that is, to Jeff and the inquisitor—that what’s at issue is whether or not Odile accepts ‘that one [Odile points at the picture] is dead’. Given that the question the inquisitor asked was ‘Doesn’t she realize that Twain’s dead?’ it is quite appropriate to use ‘Twain’ as a representative of ‘that one’.

The right answer to the question—How does e in

Odile believes that . . . e . . .

come to represent some expression (or class thereof) in Odile’s belief mediators?—seems to go something like this. The interests and intentions of a speaker (and, to some extent his audience) determine how expressions in a ‘that’-clause may be and are used to represent mediators of belief. Sometimes, for example, the speaker and audience are focusing on a specific name or term or other way of representing something. This is what is going on in the case of
Mutt and Jeff. This produces a restriction on what the expressions in a 'that'-clause can represent. In Mutt's case, a restriction like—use 'Twain' to represent 'Twain'—is operative. In Jeff's, something like—'Twain' is to represent only terms Odile associates with her current perceptual experience of Twain's picture—is operative.

Of course, in some situations we are not focusing on how someone thinks about the objects and properties about which they have beliefs. In some contexts, as the Russellian is fond of pointing out, we just don't care about the how of Odile's belief—that is, about the sentences whose acceptance constitute her beliefs—but only the Russellian what. In these situations there's no restriction operative on what 'Twain' may represent—as long as it represents a name of Twain.

Let's see if we can parley these remarks into an account of the overall semantics of belief ascription. Perhaps the most straightforward way to proceed makes use of hybrids which come from fusing sentences and Russellian referents. Think of a sentence as being a set-theoretic entity. Identify, for example, 'Twain is dead' with

(4) <'is dead', 'Twain'>.

A pure Russellian can identify the proposition a sentence expresses (in a context) with the result of replacing the expressions in a sentence with their Russellian referents. In the case of (4), this yields

(5) <being dead, Twain>.

Consider what we get if we pair off the constituents of a sentence with their Russellian interpretations. If we do this with (4), for example, we get

(6) <<'is dead', being dead>, <'Twain', Twain>>.

Such hybrids are not Russellian propositions. They are not Fregean thoughts. They are fusions of things which represent—in this case, the expressions in a 'that'-clause—with their Russellian interpretations. Perhaps we should give them a new name, since they are somewhat different from run of the mill propositions. Since they are obtained by annotating the matrix provided by a sentence with the Russellian interpretations of its parts, we call them RAMs, for Russellian Annotated Matrices.

In general, one gets RAMs by fusing English sentences and Russellian referents, German sentences and such referents, indeed, from fusing arbitrary things which could be words or representations with Russellian referents. Take an arbitrary set of objects which could play the roles of parts of a language or system of representations. Pair off these "vocabulary" items with the parts of a Russellian proposition, and you have a RAM. The RAM represents, in a somewhat crude way, what it is for a sentence made out of the "vocabulary" to express the proposition.

Think now of the believer. She accepts various sentences, each of which has a Russellian interpretation. Just as we can fuse the content sentence of a
t-clause with its Russellian interpretation, mating (4) and (5) to get (6), so can we do this to each of the mediators of the believer's belief. If we do this for all of the mediators of her beliefs, we end up with a set of RAMs. This set encodes all of the facts about the believer which are relevant to the truth and falsity of belief ascriptions to her. Let's call this set the believer's representational system, or RS.

When we ascribe an attitude, using, say

(7) Odile believes that Twain is dead

we seem to be saying something about the believer's RS, and not something simply about the collection of Russellian propositions believed. The remarks we made above suggest that what we are doing is saying (roughly) that the RAM our 'that'-clause determines represents one of the believer's RAMs.

For our RAM to represent one of the believer's RAMs, it seems necessary (but not always sufficient) that our RAM be related to the believer's in a certain straightforward way. Putting the matter crudely, the necessary condition is that, stripped of their linguistic parts, the two RAMs amount to the same Russellian proposition.

Let's not rest with a crude statement of the condition. Call the pairs of things in RAMs, consisting of a vocabulary item and an interpretation, annotations. So <'Twain', Twain> is an annotation, <'he', Twain> is an annotation, <'is dead', being dead> is an annotation, and so forth.

Say that a correlation is a rule (a function) which maps annotations to annotations and preserves reference: that is, if a correlation takes <a, b> to <a', b'>, then b is b'. It's often convenient to speak as if correlations just mapped expressions to expressions, leaving the fact that annotations contain references as understood. Thus, for example, I'll sometimes say things like such and such a correlation takes 'Twain' to 'Clemens' and 'Clemens' to 'Clemens'.

Take a RAM p, and a correlation f (assume f is defined for all the annotations in p). Consider what we get if we systematically replace what's in p with its image under f. For example, if we begin with the RAM p determined by

Hesperus is Phosphorus

and use the correlation

\[ f: 'Hesperus' \rightarrow 'H'; 'Phosphorus' \rightarrow 'P'; 'is' \rightarrow '=' \]

we obtain the RAM—call it q—determined by

\[ H = P. \]

When p, q, and f are related in this way—q comes from p via the correlation f—I say that p represents q under f.

The relation I mentioned above, the one that is necessary for the truth of a belief ascription
Odile believes that $S$ is that the RAM determined by the 'that'-clause represent, under some correlation or another, a RAM in Odile's RS. So, for example, our old friend

(7) Odile believes that Twain is dead

is true only if the RAM determined by the 'that'-clause

(6) $\langle \langle \text{"is dead"}, \text{being dead} \rangle, \langle \text{"Twain"}, \text{Twain} \rangle \rangle$

represents, under some correlation, a RAM in Odile's RS. As I remarked above, this is roughly equivalent to what the Russellian thinks to be necessary and sufficient for the truth of (6).

But we don't think it to be sufficient for the truth of a use of (7). As remarked above, sometimes we impose restrictions on the way an expression can be used to represent parts of the mediators of someone's belief. Some contexts contain restrictions on the functions we may use to correlate our RAM (the one determined by the 'that'-clause in an attitude ascription) and the RAMs in the subject's RS. When Mutt uttered (7), it was understood that 'Twain' was to represent 'Twain' and nothing else. So in evaluating Mutt's claim for truth, we are restricted in what correlations we can use. We can only use ones which map 'Twain' to 'Twain'. What Mutt said was true provided that his RAM ((6) above) represents one of Odile's RAMs under a 'Twain' to 'Twain' correlation.

A context, then, provides a collection of restrictions on correlations. We can think of each restriction provided by a context as containing three things: A person $u$, an annotation $a$, and a collection of annotations $S$. For example, Mutt's context provides the restriction

Odile; $\langle \text{"Twain"}, \text{Twain} \rangle$; $\{ \langle \text{"Twain"}, \text{Twain} \rangle \}$

A restriction involving $u$, $a$, and $S$ tells us that, in evaluating an ascription of attitude to $u$, we are restricted to using correlations which map $a$ to something in $S$. Mutt's restriction, for example, tells us that in evaluating an ascription to Odile, we are restricted to using correlations which map 'Twain' to 'Twain'.

Our remarks above may now be codified as follows: Taken in a context, call it $c$, an ascription of the form of

$a$ believes that $S$

is true if and only if the RAM determined (in $c$) by $S$ represents a RAM in the representational system of what $a$ names (in $c$), under a correlation which obeys all the restrictions operative in $c$. Or, a bit more loosely: the ascription is true if the RAM that $S$ represents one of $a$'s RAMs, given the context's restrictions on correlations.5

It should be tolerably clear how this proposal works in the case of Mutt, Jeff, and Odile. In the case of Hammurabi, too, matters seem relatively straightforward. Suppose we have heard the story of the ancient Babylonians, and say
that Hammurabi believes that Hesperus is Hesperus, but not that Hesperus is Phosphorus. Given our focus on the way Hammurabi thought about the planets, it is natural to suppose that we are operating under the restrictions

\[
\text{Ham; } \langle \text{'Hesperus'}, \text{Hesperus} \rangle; \{\text{<the Babylonian word which 'Hesperus' conventionally translates, Venus>}\}
\]

\[
\text{Ham; } \langle \text{'Phosphorus'}, \text{Phosphorus} \rangle; \{\text{<the Babylonian word which 'Phosphorus' conventionally translates, Venus>}\}.
\]

We can abbreviate here, and represent the restrictions thus:

\[
\text{Ham: 'Hesperus' } \rightarrow \text{ the Babylonian it conventionally translates}
\]

\[
\text{Ham: 'Phosphorus' } \rightarrow \text{ the Babylonian it conventionally translates.}
\]

Given these restrictions, as well as the noncontroversial assumptions—'Hesperus' conventionally translates 'H', and not 'P'; 'Phosphorus' conventionally translates 'P', and not 'H'; the facts about what Ham accepted as given at the beginning of this section—it is clear that the RAM determined by 'Hesperus is Hesperus' represents one of Ham's RAMs, relative to the context's restrictions, but that determined by 'Hesperus is Phosphorus' does not. If we are ascribing belief to Hammurabi, the latter RAM

\[
\langle \langle \text{'is'}, \text{Identity} \rangle, \langle \langle \text{Hesperus'}, \text{Venus} \rangle, \langle \text{Phosphorus'}, \text{Venus} \rangle \rangle
\]

can, given the restrictions, only represent a RAM of the form

\[
\langle \langle e, \text{Identity} \rangle, \langle \langle \text{'H'}, \text{Venus} \rangle, \langle \text{'P'}, \text{Venus} \rangle \rangle
\]

where \(H\) and \(P\) are fixed as above. (Since there are no restrictions on 'is', \(e\) can be any piece of vocabulary). Since such a RAM is clearly not in Hammurabi's RS, the claim that he believes that Hesperus is Phosphorus is false.

Isn't this a sort of closet Fregeanism, with expressions in 'that'-clauses doing duty for the senses of expressions for those to whom we ascribe attitudes? I don't want to quibble about the epithet (or honorific, depending on your perspective) 'Fregean'. If you're willing to accept this account, and it makes you feel better about things to call it Fregean, by all means do so. But we have wandered quite far from any traditional sort of Fregeanism. For example, as I observed above, this view abandons the Fregean view, that attitude ascription involves a match of nonreferential, cognitive content between a t-clause and some state of, or sentence accepted by, the subject of the ascription. Whether or not Mutt and Odile associate similar ways of thinking of Twain with their uses of 'Twain' is wholly irrelevant to the question, "Does, or could, Mutt use 'Twain' to represent Odile's uses of "Twain"?" In fact, one can comfortably hold the sort of view I am urging and insist that in general, interpersonal comparisons of a sentence's sense or cognitive role cannot be made. One could say, for instance, that such comparisons presuppose an isomorphism of the subjective probability functions users associate with their
sentences, a sort of isomorphism which, practically speaking, is never to be found.

Furthermore, there is nothing in this view which suggests that associated with a use of an expression, as a matter of the expression’s meaning or otherwise, is a descriptive condition which determines the expression’s reference. This view does not retain even the shade of the idea that sense determines reference.

It is also worth comparing the way this and Fregean views treat quantification into attitude ascriptions. For Fregeans, quantification into an attitude ascription, as in

(8) There's an \( x \) such that Odile believes that \( x \) is dead

involves implicit quantification over senses, with (8) being glossed as

For some \( x \) and \( s \): \( s \) is a sense which presents \( x \) and Odile believes \( s \) is dead\(^6\),

the “\( s \)'s here being “sense quotes.” The view I've glossed allows a straightforward, objectual treatment of quantification in. Relative to an assignment \( \nu \) to the variables, the sentence ‘\( x \) is dead’, \( x \) a variable, determines the RAM

\[ <<\text{'is dead'}, \text{being dead}>, <x', \nu (x')>>. \]

For example, if \( \nu(x') \) is Twain, then ‘\( x \) is dead’ determines the RAM

\[ <<\text{'is dead'}, \text{being dead}>, <x', \text{Twain}>>. \]

This can represent one of Odile's RAMs just as well as the RAM that Twain is dead. So quantification in is treated quite straightforwardly, with (8) true just in case ‘Odile believes that \( x \) is dead’ is true for some assignment to \( x' \).\(^6\)

It would be misleading to say that the view of attitude ascriptions I am sketching has nothing in common with Fregean views. Like the Fregean, I see the truth of an attitude ascription as being sensitive to facts about the way an individual represents objects and properties. But I reject the mechanism the Fregean proposes to explain this sensitivity. I also would prefer to be committed to as “thin” an account of ways of thinking as possible. I am somewhat dubious of the significance of comparisons of nonreferential content across speakers, and am trying to offer an account which avoids them. In fact, so far I've said nothing that commits us to a very interesting notion of intrapersonal identity of nonreferential content, between different expressions or “modes of thought.”

My account also has a good deal in common with Russellian accounts of attitude ascription. It rejects the idea that senses or nonreferential, cognitive “contents” like conceptual roles are apt semantic values for ‘that’-clauses. It is compatible with views about reference—for example, that names and demonstratives are directly referential—which partially motivate Russellianism. My view honors the intuitive evidence for Russellianism—for example, that the truth of an attitude ascription is often unaffected by substituting one
directly referential expression for another which names the same thing. It
gives a natural, objectual account of quantifying in. But unlike Russelian-
ism, this account is consistent with the idea that, for example, someone might
know that Clark Kent is Clark Kent, but not know that Clark Kent is Superman.

II

On my view, a 'that'-clause of English names something quite different from
its natural translation into German. ‘that Twain is dead’ names something in
which the expression ‘is dead’ occurs; its German translation, ‘dass Twain tot
ist’ names something in which the German expression ‘ist tot’ occurs. You
might anticipate that this would make English belief ascriptions and their nat-
ural German translations diverge in truth conditions in an objectionable way.

This, in essence, was Church's complaint against Carnap. I do not
think that any version of this objection has force against the account I've
given. Consider first

(1) Odile believes that Twain is dead

and its German translation

(1') Odile glaubt, dass Twain tot ist.

If a sentence has indexical elements, we can speak of its truth conditions only
relative to a particular context or use. We are treating ‘believes’ as an indexi-
cal, and would give its German translation ‘glaubt’ a perfectly parallel treat-
ment. So both (1) and (1') can be assigned truth conditions only relative to a
context.

Now, there are a number of worries you might have about the way in
which truth conditions are assigned to uses of (1) and (1') on this theory. For
example, you might complain that a use of (1) in one context and a use of (1')
in some other context could have different truth conditions. And this, you
might say, is unacceptable, (1) and (1') being translations of one another.

Of course (1) and (1') can be used so that they have different truth con-
ditions. But this is no objection: One of the motivations for this view, you will
recall, is that sentence (1) itself can be used in such a way that it says different
things in different contexts. One reason for adopting the view is that a sen-
tence like (1) is used some, but not all, of the time to say something quite
specific about the how of Odile's belief. The same is true of (1'). So the obser-
vation that (1) and (1') might be used to say different things is not an objec-
tion to the theory.

For another thing, you might worry that no use of (1) can have precisely
the truth conditions of a use of (1'). If this were true it would be a serious
objection.

However, consider ‘believes’ and ‘glaubt’. There is no reason that they
can't have exactly the same meaning. (To use Kaplan's terminology: They have
the same character.) The verbs' common meaning is something which given a contextually supplied "translation manual" (i.e., a set of restrictions) pairs off people and RAMs: Person \( P \) is paired with a RAM \( R \), provided \( R \) represents, under the manual, a RAM that \( P \) has in his RS. Of course, the manuals the English verb is given are usually different from those the German verb gets. So in context the verbs typically have different contents. But this is consistent with their having the same meaning. (The situation is to some extent parallel to that of 'I' and 'ich'—in contexts they may refer to different things; the two still have the same meaning.)

Thus it is perfectly possible for uses of (1) and (1') to have the same truth conditions. (I suppose truth conditions to be sets of worlds.) If (1) is used in a context in which no restrictions are operative, then (speaking roughly) (1) says that Odile accepts some sentence of the form \( D(t) \), where \( D \) refers to being dead, and \( t \) refers to Twain. Used in a context without restrictions, (1') says exactly the same thing.

Even when restrictions are operative, (1) and (1') may have the same truth conditions. For example, suppose (1) is used in a context \( c \) in which the restrictions require that 'is dead' be mapped to 'is dead' (and no other restriction is operative). A use of (1') in a context in which just the restriction Map 'ist tot' to 'is dead' is operative will be one in which (1') is assigned the same truth conditions as is (1) in \( c \).

Here is another issue raised by Church's objections to Carnap's account of attitude ascriptions. Church himself noted that his objection to Carnap was really only forceful for the case of iterated attitude ascriptions like

\[ (2) \text{ John believes that Hammurabi believes that Hes is hot} \]

and their natural translations into foreign languages.\(^8\)

What will we assign a 'that'-clause in which 'believes' occurs, like 'that Hammurabi believes that Hes is hot'? What we need to know is whether, in constructing a RAM for this, we pair off 'believes' with its meaning (the rule for getting from context to content) or with its content in a particular context. It's clear that for most indexicals we should pair the indexical off with its content. Thus, for example, in a context where I am speaking, the 'that'-clause in

\[ (3) \text{ Peter believes that I am fat} \]

should name the RAM

\[ (3) \langle'\text{is fat}', \text{being \textit{fat}}, <'I', \text{mark richard}>\].

The exception to this rule is the verbs of attitude. In constructing the RAM that Hammurabi believes that Hes is hot, 'believes' is paired with its (context-independent) meaning. We also think of the RAMs in the believer's representational system, which correspond to sentences in which 'believes' occurs, as being constructed out of the meaning of 'believes', not out of its content in the believer's context.
If we proceed in this way, there’s nothing particularly puzzling about iterations of ‘believes’. Consider, for example, (2) and its natural German translation. The only difference between what the English

that Hammurabi believes that Hes is hot

names and what its natural German translation names is in the vocabulary items in the RAMs. There’s no difference in the semantic values paired with the vocabulary.

So what we will say here will be exactly parallel with what we would say about (1) and (1‘). For example, suppose that we use (2) in a context without restrictions. Suppose further that John accepts the sentence ‘Hammurabi believes that Hes is hot’. Then what we say is true. It’s easy to work out that a use of the natural German translation of (2) (‘John glaubt, dass Hammurabi glaubt, dass Hes heiss ist’, or something of the sort), used in a context with no restrictions, will be true in the same circumstance. In fact, it will have the same truth conditions.

A last, related objection. Ali Kazmi pointed out to me that on my view, a sentence like ‘Odile believes that Cologne is large’ and its translation into a foreign language (say, ‘Odile glaubt, dass Koln gross ist’) may differ in truth value taken relative to one context. For the context may provide restrictions which deal only with English vocabulary, not with foreign language vocabulary. Kazmi suggested that this was a problem.

If it is a problem, it is because of the truth of a principle like:

(I) If a sentence type S of one language is naturally and correctly translated by a sentence type T of a second language, then there is no context relative to which S and T do not have the same truth values.

I doubt that (I) is true. As I understand it, there is a sort of animal, the woodchuck, for which English has two expressions (‘woodchuck’ and ‘groundhog’), where French has but one (‘la marmotte’). The French ‘Louis croit que Chuck est une marmotte’ is thus correctly and naturally translated by both ‘Louis believes that Chuck is a woodchuck’ and ‘Louis believes that Chuck is a groundhog’. But presumably, the latter two sentences can diverge in truth value relative to some context. So the French and a natural and correct translation thereof will diverge in truth value relative to some context.

III

I want to discuss Kripke’s puzzle about belief, both in its version involving Pierre, ‘London’, and ‘Londres’, and that involving Peter and ‘Paderewski’. I assume familiarity with the case of Pierre. Kripke writes “This is the puzzle: Does Pierre, or does he not, believe that London is pretty?” Kripke argues successively that

(1) Pierre believes that London is pretty
is (or at least seems) true (focusing, of course, on Pierre’s “French beliefs”); that

(2) Pierre believes that London is not pretty

is (or at least seems) true (focusing now on Pierre’s “English beliefs”); but that they can’t both be true, because Pierre, a leading logician, “would never let contradictory beliefs pass. And surely anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them.”

What counts as a solution to the puzzle? Certainly, all else being equal, we want to preserve as much of our pre-theoretic intuitions about sentences (1) and (2) as possible. This suggests that in the ideal, a solution to the puzzle will provide a way of saying that both sentences are true—that is, the uses Kripke makes of them, when he ascribes truth to them, are true uses; and yet (1) and (2) together can’t be true, for more or less the very reason Kripke gives.

One nice thing about the view I’ve been sketching is that it does allow us to say something like this. One way in which restrictions on correlations should come to be operative is this: If the speaker is focusing on how someone expresses his beliefs, thinks that his audience is so focusing (and thinks his audience thinks he knows that they are so focusing), then the appropriate restrictions tend to come into play. So when Kripke begins walking us through the puzzle, focusing on Pierre’s French beliefs, we might expect the restriction

(3) ‘London’ $\rightarrow$ ‘Londres’

to be operative. This will make (1) true. Then Kripke asks us to focus on Pierre’s English language beliefs. The old restriction is no longer operative, and a new one

(4) ‘London’ $\rightarrow$ ‘London’

comes into play. So (2) is true. Then Kripke asks us, in effect, to answer the question: Can (1) and (2) be true together? And he observes that they seem to imply that Pierre is in some sense irrational.

And this is quite correct. Any natural way of evaluating (1) and (2) in the context of one conversation will use a single correlation function. It would be unnatural to adopt a completely new correlation each time we reattribute belief (or desire or another attitude) to an individual within the course of a single discourse. So any natural way of evaluating the conjunction of (1) and (2) makes it false. And it is false for just the reason Kripke suggests: It suggests that Pierre is irrational, accepting some sentence and its negation. It is (roughly) in this sense that anyone with contradictory beliefs is in a position to recognize such and correct it.

The case of Peter and ‘Paderewski’ may appear to present something of a problem for my view. The case seems like that of Pierre and ‘London’-‘Londres’: Peter hears one day of a famous musician, Paderewski, and thinks
to himself 'Paderewski had musical talent'. He hears on some other day of a Polish statesman, Paderewski. Thinking that politicians are poor musicians, he thinks to himself 'Paderewski did not have musical talent'. We have, Kripke urges, exactly the same sort of puzzle about

(5) Peter believes that Paderewski had musical talent

(6) Peter believes that Paderewski did not have musical talent

as we did concerning the Pierre sentences.

I suggested that on any natural reading of the Pierre sentences they wouldn't be true in the same context. It would seem that I can't say this sort of thing about this case. For here there seems to be but one sentence (type), that of

Paderewski had musical talent

one such that Peter accepts both it and its negation. So it seems that no matter what restrictions are operative, if (5) is true, so is (6). So, it appears, I have to treat like cases (Pierre and Peter) in unlike ways.

I do not think I do. I could say that the expression 'Paderewski' in Peter's spoken dialect is ambiguous: appearances to the contrary, Peter does not accept some natural language sentence and its negation. I think this somewhat ad hoc, so I will try to get by without saying it.

Let's go back to the picture of belief that we are working with. Think of the believer as having sentence tokens written inside of his head, on a blackboard next to the pineal gland. Think of his representational system, the set of RAMs we quantify over in giving truth conditions for attitude ascriptions, as being determined somehow by what's written on the blackboard.

Up to now, I have been pretending that the way the blackboard determined the RAMs was as follows. Suppose a token of 'Twain is dead' is on Odile's blackboard. Then a RAM that looks thus:

<<'is dead', being dead>, <'Twain', Twain>>

is in Odile's RS. Here, the quoted items are the types of the relevant tokens. On this picture, we would have in the case of Peter a RAM and its negation in Peter's representational system. (I haven't defined a negation-of operation for RAMs, but it's obvious, I think, how to go about it.)

I want now to distinguish between the sentence tokens on the blackboard and representations. I think of the sentences on the believer's blackboard as determining representations; in constructing RAMs to put in the believer's RS, we form them from the representations which the sentences on the blackboard determine. Instead of putting expression and referent in the RAM, we pair the representation determined by the expression with the expression's referent.

To return to the example above: If Odile tokens 'Twain is dead', then her RS contains the RAM
where \( R_1 \) and \( R_2 \) are the representations determined by her tokens of 'is dead' and 'Twain', respectively.

We could identify representations with sets of tokens on the blackboard. Then we have only the problem of saying when two inscriptions on the blackboard determine the same representation. I won't give a full blown theory about this. But I think such a theory is possible, and I think its broad outlines are fairly clear.

As I see it, there will be two sorts of conditions which together will be necessary and sufficient for two tokens to determine the same representation. Intuitively, the two sorts of conditions are what we could call "outside" and "inside" conditions. The outside conditions will be broadly causal in nature. If we have two proper name tokens written on the blackboard, for example, they will have to be residues of the same causal or historical chain (and thus refer to the same thing) in order to satisfy the outside condition. I take it that this will result, if we make the outside condition (in the case of proper name tokens) that two proper name tokens determine the same representation if and only if they are tokens of the same name (type). Let us do this. Observe that this rules out, for example, 'Hesperus' and 'Phosphorus' determining the same representation.

The interior condition should be a "recognition" condition. Let me speak rather intuitively about this. Usually, when we hear someone talking about someone, we think we know who is being talked about. We hear someone say 'Reagan is going to bomb Nicaragua', and assume that it's Reagan the president who is being discussed, not the Regan the animal rights philosopher. When this happens, we somehow "file" the token of 'Reagan' we are hearing with certain other tokens ("presidential tokens") on our blackboard, and segregate if from others. In such a case, the new token of 'Reagan' has the interior relation to the presidential tokens. Our segregating the new token with the older ones is a sort of "recognition." Suppose that it was the president that was being talked about, and that outside conditions were satisfied. Then, with both outside and inside conditions satisfied, the new token of 'Reagan' and the older presidential tokens will all determine the same representation.\(^{12}\)

I haven't given a theory of representations here, of course. At the least, before I have even the vague outlines of a theory, I need to talk about demonstratives and sensuous experiences. It's a lacuna, but for now I have to live with it.

In the case of Peter, we have a token of 'Padereweski has talent' on one place on the blackboard, a token 'Padereweski has no talent' somewhere else. It is quite clear, I hope, that on any way of spelling them out, the interior conditions will not be satisfied with respect to the two tokens of 'Padereweski'. So they determine different representations.

This means that, in principle, we can treat the Padereweski case just as we treated the case of Pierre.
IV

I close by mentioning some issues of a broadly logical nature. First of all, we can introduce propositional quantifiers and variables, letting them range over the class of all RAMs. These are treated as one would expect, with

For all \( p \): \( a \) believes \( p \) \( \rightarrow \) \( b \) believes \( p \)

being true just in case any assignment of a RAM to \( 'p' \) makes the matrix sentence true.

A second issue is that of RAMs and truth bearers. Traditional Russellian propositions are both the objects of attitudes and the bearers of truth. Many RAMs are indistinguishable, \( qua \) truth bearers, from such. The RAM of 'Hesperus is hot', for example, is more or less the Russellian proposition that Hesperus is hot with some extra material in it, material we can ignore for the purposes of assigning truth conditions.

But RAMs corresponding to sentences in which belief predicates occur are in a certain sense incomplete. For these RAMs contain the (context constant:) meaning of 'believes', not one of its contextually varying contents. If we appoint RAMs to the office of official bearers of truth, then we will have some truth bearers—e.g., that Odile believes that Twain is dead—which are true in some contexts and false in others.

This means I'm committed to either (a) contextual variation in the truth value of the bearers of truth, or (b) uses of a sentence like 'It's true that Odile believes that Twain is dead' ascribing a property not to the RAM that Odile believes that Twain is dead, but to a supplemented RAM, a pair consisting of the RAM and the restrictions operative in the context of use.

There's nothing objectionable \( per se \) about either (a) or (b). There is, after all, a long tradition which sees the object of belief as changing truth value across contexts. Some have wanted to say, for example, that the proposition that Reagan is president is true now, but—patience—won't always be true. I happen to disagree that the objects of belief change truth value in response to change of time. But just because objects of belief don't change truth value in response to some contextual variations doesn't mean that they don't respond to others. And there is evidence that certain objects of belief have their truth value relativized, in one way or another, to context. When Mutt and Jeff say 'Odile thinks that Twain is dead', we agree that, within the confines of their respective conversations, one may speak truly, the other falsely. But they both believe that Odile believes that Twain is dead. So it looks as if one thing, their common belief, is true at one place, false at another.

In any case, we evaluate beliefs for truth, as well as tie the truth of beliefs to the truth of the sentences which express them. So we should give some account of how truth and RAMs are related. A natural way to do this is as follows. Let us say that a supplemented RAM is a pair consisting of a regular RAM and a set of restrictions. Note that a sentence taken in a context
determines a unique supplemented RAM: pair the RAM the sentence determines with the contextual restrictions.

A supplemented RAM will determine truth conditions, and thus a truth value at the actual world, in a straightforward way. If the RAM corresponds to a sentence with no belief operator, the supplemented RAM has the truth conditions of the corresponding Russellian proposition. If the RAM corresponds to a sentence with a belief operator, one uses the set of restrictions in the supplemented RAM to turn the meaning of the 'believes' operator into a content, and then assigns truth conditions in the way suggested by the truth definition for belief ascriptions.

We can thus introduce a predicate 'is true' which combines with 'that'-terms to form sentences. "That S is true" is itself true in a context c iff the supplemented RAM consisting of the RAM named by 'that S' in c and the set of restrictions in c is true. The predicate behaves nicely enough. We have such intuitively satisfying results as the validity of the argument schema:

For all p: a believes p \rightarrow p is true
a believes that S
Thus, that S is true. 13

Finally, I'll say something about reference to, and individuation of, propositions. We ought to treat 'that S' and 'the proposition that S' in the same way. So let terms of the latter form name what those of the former do.

People will object that this cuts propositions far too finely. I can't answer every objection to the fineness of the cut here, so I'll address only the one I think most serious.

Objection: Surely the sentences 'snow is white' and 'la neige est blanche' say the same thing, a thing which Odile believes. So there's something, p, such that Odile believes p, 'snow is white' says p, and 'la neige est blanche' says p. You have to deny this, because you say that the proposition that snow is white is distinct from that named by 'que la neige est blanche'.

Response: Well, this depends upon the exact semantics of 'say' as it is used in "'snow is white' says that snow is white," doesn't it?
Note, in working up to an account of this use of 'says', that it does not make sense to assess what a sentence says, unless we imagine the sentence taken relative to some context or other. After all, 'London is pretty' says one thing at noon (that London is pretty then), and another at six o'clock.

Suppose we take a sentence relative to a context c. The sentence will determine a RAM p in the context. We can ask whether p represents, relative to the restrictions of the context, various other RAMs. Call the set of RAMs so represented p's profile in the context. (Strictly, the profile of a RAM is a function from worlds to pairs of an individual and a RAM, since restrictions are keyed to individuals.) Then we might say that

"S" in context d says that T
is true in a context \( c \) iff the profile in \( c \) of the RAM determined by \( T \) in \( c \) is identical with the profile in \( d \) of the RAM determined by \( S \) in \( d \).

If we suppose that we are working in a context in which no restrictions are operative, and that ‘la neige est blanche’ and ‘snow is white’ are to be taken relative to that context, then it will be true that ‘snow is white’ says (in the context) that snow is white, and ‘la neige est blanche’ (in the context) says that snow is white. And so it will be true (assuming Odile’s beliefs are rightly arranged) that there’s a \( p \) such that Odile believes \( p \), ‘snow is white’ says \( p \), and ‘la neige est blanche’ says \( p \).14

Notes

1. I defend this assumption in chapter 1 of *Propositional Attitudes* (Cambridge, forthcoming).
2. Such a view is strongly suggested, for example, by some passages in Kaplan’s “Demonstratives,” in *Themes From Kaplan* edited by Almog, Perry, and Wettstein (Oxford, 1988). Nathan Salmon invokes a more general notion, that of being acquainted with a proposition under a guise or way of apprehending such. (See Frege’s Puzzle [Cambridge, Mass., 1986].) Salmon’s guises are sententially structured at least to the extent of having parts corresponding to constituents of the proposition: “The means by which one is acquainted with a singular proposition includes as a part the means by which one is familiar with the individual constituent(s) of the proposition” (108).
3. I have argued for this in “Taking the Fregean Seriously,” in *Philosophical Analysis: A Defense by Example*, edited by David Austin (Dordrecht, 1988). See also chapter 2 of *Propositional Attitudes*. Kripke also notes this problem: see Kripke’s “A Puzzle about Belief,” in *Meaning and Use*, edited by A. Margalit (Dordrecht, 1979).
4. I ignore, here and elsewhere, complications introduced by tense.
5. It is probably not obvious what the semantic value of the verb ‘believes’ is supposed to be. Simplifying somewhat: ‘Believes’ and its ilk are to be treated as indexicals. They have a constant meaning, or, to use Kaplan’s term, character; their interpretation, or content, varies from context to context. The character of ‘believes’ is a rule which, given a collection of restrictions, returns an appropriate intension. Since ‘believes’ looks at a person and a RAM—it acts as a dyadic predicate which joins a term like ‘Odile’ with something that names a RAM, like ‘that Twain is dead’—appropriate intensions for ‘believes’ will be functions from pairs, of individuals and RAMs, to sets of worlds. Write

\[
\text{Rep}(p,q,f)
\]

for ‘\( p \) represents \( q \) under \( f \)’. Write

\[
\text{Obey}(r,f,u)
\]

for ‘\( f \) obeys all the restrictions in \( r \) which are relevant to \( u \)’. (A restriction \(<x,a,S>\) is relevant to \( u \) iff \( x = u \); \( f \) obeys the restriction if \( f(a) \) is in \( S \).) Then, if \( r \) is the collection of restrictions operative in \( c \), ‘believes’ takes as value in \( c \) the function which maps \(<u,p>\) to the set \( X \) of worlds such that \( w \) is in \( X \) iff, at \( w \), \( u \) has a RAM in her RS such that, for some \( f \), \( \text{Obey}(r,f,u) \) and \( \text{Rep}(p,q,f) \).
The simplification here is that of making 'believes' two-placed. In *Propositional Attitudes*, the proposal of this paper is given an alternative development, in which the verb is made three-place, with the extra argument place one for correlations. This allows the nicest formalization of the treatment of Kripke's puzzles which is given in section III of this paper. The resulting system retains the virtues of the system of this paper—e.g., it avoids Church-style objections and allows the sorts of generalizations discussed in section IV.

6. I assume that a context cannot provide restrictions on how free variables are treated by correlations. This is necessary (and sufficient) to validate arguments like Odile thinks that Twain is dead. Thus, for some x, Odile thinks that x is dead.


8. Ascriptions like (2) raise the following problem. Consider an ascription like (2) in the context of a pure Russellian account. The Russellian assigns a relation to 'believes' and a structured entity to the 'that'-clause. The assignment to the 'that'-clause contains the semantic values of the expressions in the 'that'-clause. So if the 'that'-clause contains 'believes'—as it does in (2)—we have the relation assigned to the main verb—the belief relation—trying to relate something which contains that very relation. Besides sounding like the opening of a book by Kierkegaard, this is more or less like a function taking itself as an argument, something which on the normal set-theoretic understanding of a function is impossible.

I have investigated what happens, in a system like the one I have been sketching, if we stratify the belief predicate, more or less in the way Tarski stratified the truth predicate, by introducing a series of languages. (This requires that RAMs come in levels, as well.) I have found nothing untoward in the system, save that it requires fragmenting 'believes' (as Tarski's approach, adapted to natural language, would require fragmenting 'true'.)

I am aware that many people object to this way of treating 'true' in natural languages. I believe that other treatments of 'believes' would also work here—in particular, I think that modeling propositions and the semantic values of expressions like 'believes' in a set theory like that of Peter Aczel's, in which sets can be members of themselves, would mesh quite well with the approach I am sketching, and would yield a more satisfactory treatment of 'believes'. (See Aczel's *Nonwellfounded Sets*, CSLI Lecture Notes, no. 14, [Stanford, Calif., 1988].) But at present I have not fully investigated this.

In the text, I suppress all reference to this complication. This paper is about the way expressions—in particular, singular terms and garden variety predicates—function within the scope of attitude verbs. It is not about the semantical paradoxes or paradoxes like the knower. While the latter are of course important, I think it is fair to ignore them in discussing the current problems. One insoluble problem at a time, please.

In section IV, when I discuss the introduction of a truth predicate, I again ignore the fact that some method to forestall paradox is needed. Again, I have investigated a series of languages approach here, each with a truth predicate for the languages below, in the context of implementing the proposal of this paper. Other than objections to the levels approach, which are not relevant to the issues this paper is concerned with, I do not see any problems with the resulting system.

10. Conversational principles (“Don’t be confusing”) alone dictate this much. Furthermore, only if we assume that sets of ascriptions are to be so interpreted does the widespread belief—that explanations of action in terms of beliefs and desires are at least candidates for true explanations—even make sense. For example, the claims

Randi thought that if there were waving, Ann would smile

Randi wished that Ann smile

evaluated together, instead of separately, are true only if (roughly speaking) part of Randi’s overall text (the family of sets of sentences believed-true, wished-true, and so on) looked thus:

Belief: $W \rightarrow Sa$ ; Desire: $Sa$

where $W$, $s$, and $a$ are expressions determining the set of worlds in which there’s waving, the property of smiling, and Ann, respectively. But it is plausible to think that this being true is a reason for thinking that Randi waved. So the explanation which goes ‘Randi waved, because he thought that if there were waving, Ann would smile, and he wanted Ann to smile’ can plausibly be said, in a very straightforward sense, to be a genuine explanation. If, on the other hand, we switch correlations in midstream, then the truth of the belief and desire ascriptions gives us no reason to think that Randi will wave.

The fact that on this view we can make sense of the idea that action could be explained by attitude ascription is surely a strong reason to prefer this approach to Russellianism. For on Russellian grounds the truth of the premises of a belief-desire explanation give one absolutely no reason whatsoever to think its conclusion true.

It is not trivial to systematically assign truth conditions to complex sentences (i.e., ones containing truth functional and temporal operators, as well as quantifiers), once we allow that sentences of the form he believes that $S$, and he believes that $T$ can have truth conditions on which they imply that that $S$ and that $T$ represent RAMs under a single correlation, especially if we want to preserve the validity of classical logical principles. I try to characterize one way of assigning truth conditions, while preserving classical logic, in chapter 4 of Propositional Attitudes.

11. I do have to say that there is a way of interpreting the conjunction of (2) and (3) so that it is true. If we are in a context in which no restrictions are operative, and we interpret the ascriptions singly, and not jointly, then both (2) and (3) will be true. I don’t think that this is a disaster. First of all, as I argued above, this is a very unnatural way to evaluate multiple ascriptions. So, I think, I may fairly say that the account I’ve been presenting leads us to expect that there will be at best a weak inclination among speakers to say that (2) and (3) are both true, if they are presented together. And this is what we do seem to find.

And we do have some inclination to say that in the case of Pierre that both are true. Most of us are inclined to reason as follows: Well, (2) is true. And, gosh, (3) is true. (Notice one does this by evaluating (2) and (3) individually.) So I guess (gulp) their conjunction is true.

12. I assume that all the old presidential tokens are tokens which refer to Reagan the president. Incidentally, I take the interior, “recognition,” condition to be such that terms which do not name the same thing may satisfy it. For example, suppose I hear ‘Reagan will bomb Nicaragua’, and see someone point at a man and say ‘he is president’. If I “group” ‘Reagan’ and ‘he’ together (I take the same person to be spoken of and ostended), the terms satisfy the interior condition, even if the man pointed at was not Reagan.
I should say here that, while none of what I say can be attributed to David Kaplan, much of it is influenced by various remarks he has made about identity of words, recognition, and other matters.

13. I suppose consequence to be defined as follows: Sentence \( A \) is a consequence of a set \( S \) of sentences (in a model \( M \), relative to a context \( c \) thereof) iff for any \( w \) in the model, if \( w \) is in the intension of each member \( S \), taken relative to \( c \), then \( w \) is in the intension of \( A \), taken relative to \( c \). (Intuitively, this amounts to: \( A \) is a consequence of \( S \) iff, taken in any context, what \( S \)'s members say jointly entails what \( A \) says.) Logical consequence is a relation between sets of sentence types defined by the obvious generalization of the above.

14. I assume, of course, that 'la neige' is a syntactical unit.

I am indebted to Ed Gettier for a discussion of an earlier draft of this paper, as well as for conversations over the years on the topics of this paper. David Austin, Jody Azzouni, Harold Levin, Graeme Forbes, and Ali Kazmi made helpful editorial and philosophical comments on versions of the paper, for which I thank them as well. I acknowledge a general debt to Richard Grandy, who tried many years ago to convince me that attitude ascription had something to do with translation.