MARK RICHARD

DIRECT REFERENCE AND ASCRIPTIONS OF BELIEF

It is often supposed that demonstratives and indexicals are *devices of direct reference* – that they are, as David Kaplan puts it, terms which "refer directly without the mediation of a Fregean *Sinn* [or individual concept, set of properties, etc.] as meaning."[1] Most of the resistance to this view, I think, arises from the suspicion that it is not possible to give an acceptable treatment of the semantics of belief ascriptions and other so-called propositional attitude contexts which is consistent with the thesis of direct reference. For it seems that a straightforward construal of the thesis (along with some plausible semantical assumptions) requires that demonstratives, when co-referential, be intersubstitutable everywhere, even in belief contexts, *salva veritate*. But many feel that it is *obvious* that such substitutions do not always preserve truth.

The purpose of this paper is to motivate and present a semantics for a first-order treatment of belief ascriptions which is both consistent with the thesis of direct reference and intuitively satisfactory. The paper is structured as follows: In Section I, I discuss semantical consequences of the thesis of direct reference – in particular, what it does and does not require with respect to the overall form of a semantical treatment of belief ascriptions. I also discuss a view about belief, championed by Kaplan and John Perry, which I call the triadic view of belief. Crudely put, it is the view that belief is a triadic relation among a person, a proposition, and a sentential meaning, the latter entity a different sort of thing than a proposition. On this view, to believe a proposition is to do so 'under' a sentential meaning.

The champions of the triadic view of belief have shied away from using the view to motivate a semantic account of belief ascriptions.[2] But the triadic view of belief suggests, as I note at the end of Section I, that ascriptions of belief not only imply that a proposition is an object of belief, but that it is believed in a certain way. The purpose of Sections II and III of the paper is to show that an account of belief ascriptions, on which they behave in just this way, can be formalized rather easily, and that it nicely handles certain cases which, at first blush, seem quite problematic for those who

Journal of Philosophical Logic 12 (1983) 425–452. 0022–3611/83/0124–0425$02.80

accept the view that demonstratives are directly referential. Section II is concerned with the semantics of ascriptions of belief *de se*: That is, with giving a first-order syntax and semantics which adequately represents (different readings of) ascriptions of the form ⌜*a* believes himself to be *F*⌝ (and of allied forms) and the relations of such asciptions to *de re* ascriptions of belief. Section III is concerned with semantics for 'standard' belief ascriptions in which the sentential complement to 'believes that' contains demonstrative and indexical terms.

I

The core of the thesis that demonstratives are directly referential is negative.[3] A somewhat long-winded way of expressing the thesis is this: Associated with well-formed expressions of English (taken relative to a context) is an entity which we may call the expression's *content*. (As will become clear, I am using this expression, as well as the term 'proposition', in a very technical and circumscribed way.) Contents play a number of semantic roles. One of the things the content of a singular term does is to determine, relative to a possible circumstance of evaluation, an individual; the individual so determined at a circumstance, by the content of a term (taken relative to a context), is the referent of the term (taken relative to the context) at the circumstance.[4] To say that a term is directly referential is to say something about how its content determines an individual: It is to say that it does *not* do this by means of a complex of properties, a Fregean sense, an individual concept, etc. A picturesque way of putting the matter is this: The content of a directly referential expression, taken relative to a context, *is* that thing which the expression, taken relative to the context, has as a referent at any circumstance of evaluation.

What specific semantical consequences we suppose the thesis of direct reference to have depends, of course, upon what sort of semantical assumptions we make, beyond the assumption that demonstratives are directly referential. As noted at the beginning of this paper, it is often assumed that one consequence of the thesis of direct reference is that any ascriptions of belief which differ only in that one contains a demonstrative *d* where the other contains a demonstrative *d′* have the same truth value relative to any context in which *d* and *d′* are co-referential.

Those who think that this is a consequence of the thesis of direct reference seem to reason as follows: The content of a declarative sentence, taken relative to a context – I'll use the term *proposition* for sentential contents – is what determines the truth value of a sentence, relative to the context. Content, however, is determined functionally. More precisely: The content of a sentence in a context – what proposition it expresses – is a function solely of the contents of the parts of the sentence (in the context) and the syntax of the sentence. Now the thesis of direct reference surely requires that demonstratives that denote the same thing have the same content. Hence, given this thesis, one must say that *any* sentences, differing only with respect to co-referential demonstratives, express the same proposition and, therefore, have the same truth value.

An advocate of the view that demonstratives are directly referential *could*, I suppose, deny that co-referential directly referential terms have the same content, where the content of a term is characterized as it was above. But such a denial, I think, makes the claim, that a certain sort of expression is directly referential, extremely mysterious. For recall that the claim that a term is directly referential is the claim that the term does not have, as a content, a sense, an individual concept, etc. This, coupled with the fact that such expressions are supposed to behave as rigid designators (*viz*., a use of such a term has the same referent at every circumstance of evaluation at which it has any referent at all), makes it difficult to see how one could coherently maintain that co-designative directly referential terms can have different contents. How, one wants to know, could they differ?

A more plausible response to the above argument, I think, is to deny that it is invariably the case that the content of a complex expression is a function of the contents of its parts and their syntactic mode of combination. For there is no reason to think that content is the only sort of semantic value which expressions may have. Indeed, those who subscribe to the thesis of direct reference generally recognize at least one other sort of semantic value which expressions have. Thus, for example, Kaplan holds that the linguistic meaning of an expression (meaning in the sense of what is known by one who understands an expression) is to be identified, not with its content, but with what Kaplan calls the expression's *character*, the function which takes a context to the content of the expression therein. If the only sort of semantic value we recognized was content, it would be strange, to say the least, to suppose that the determination of content was not functional across the board. However, once we recognize yet another sort of semantic value, it is far from clear that we should suppose that the content of the whole is invariably a function of the contents of the parts.

This is particularly unobvious when the second semantic value is related to content as is character: In a fairly clear sense, character determines content, not the converse. Should the language contain an operator which is sensitive (not just to content, but) to character, it is at least an open question as to whether content is invariably functional. For suppose there were an operator, O, such that $O(A)$ is true only if the character of A has P. Since expressions with distinct characters may, relative to a context, have the same content, it is at least *a priori* possible that there be expressions B and B' such that the character of B has P, the character of B' does not, but, relative to some context c, B and B' have the same content. In this case, $O(B)$ and $O(B')$ have different contents (*viz*., express different propositions) relative to c, since they diverge in truth value.[5]

I contend that this is not just an idle possibility: In Section III, I will urge that 'believes that' is sensitive to linguistic meaning (construed as character) as well as to content. This, however, is anticipating matters somewhat. For the moment, all we have established is this: The advocate of direct reference need not assent to the view that $\ulcorner a$ believes that $S\urcorner$ and $\ulcorner a$ believes that $S'\urcorner$, taken relative to a context in which S and S' express the same proposition, invariably have the same truth value.

It is worth observing that the above argument is consistent with the claim that, in a circumscribed but nonetheless significant number of cases, the content of an expression is determined as a function of the contents of its parts. In particular, we have given no reason to think that the advocate of direct reference would deny that what proposition is expressed by a sentence which contains at most truth functional or simple intensional operators (*viz*., temporal or modal operators) is determined functionally by the contents of its parts and their mode of combination. I shall assume, for the purposes of this paper, that an advocate of the view, that demonstratives are directly referential, would say this.

We are now ready to discuss belief and ascriptions of belief. A fairly 'standard' view of belief is that propositions, characterized as above, play the role of objects of belief in two senses: (a) They are objects of belief in the sense that belief is a dyadic relation, the second term of the relation being a proposition; (b) they are objects of 'belief', in the sense that an ascription of belief $\ulcorner a$ believes that $S\urcorner$ is true iff what a denotes bears the belief relation to the proposition expressed by S.

Many who are sympathetic to the thesis of direct reference – notably Kaplan and John Perry[6] – have proposed that propositions are objects of belief in some sense, but that the relation between a person and a proposition, when the latter is an object of his belief, is somewhat more complex than the above account suggests. On this view – the triadic view of belief, as I will call it – belief is a triadic relation between a person, a sentential meaning (understood as being a Kaplanesque character), and a proposition; to believe a proposition is to do so under a sentential meaning.[7] I will use the term *acceptance* for the relation which one bears to a sentential meaning when one believes a proposition under it; it is to be understood that, on the triadic view, a proposition *p* is an object of someone's belief if and only if he accepts a meaning which, relative to the context of which he is the agent, has *p* as value.

It is not my purpose here to defend this view of belief. I will, however, note that a view of belief along these general lines seems mandatory for those who accept the thesis that demonstratives are directly referential and take propositions to be, in some important sense, objects of belief. For, to take an example, on the thesis of direct reference (making, of course, the kind of semantical assumptions we are currently making), someone who expresses something he believes by saying 'You [person *X* is addressed, say, through the telephone] are happy, but she [*X*, who is standing across the street, is demonstrated] is not happy' expresses belief in the same proposition as does one who addresses *X* and says 'You are happy, but you are not happy'. Without invoking a view like the triadic view, it is difficult to explain, or even explain away, the intuition that an irrationality is present in the latter belief which is not present in the former – for the object of belief, in the sense of proposition believed, is the same in both cases. Invoking the view, however, what one can say is this: What is irrational is not to have the proposition in question as an object of belief, but to believe it in the way the second person does. For a rational person who understands English must know that the second meaning cannot yield a truth.[8]

It is worth noting that on such an account of belief, propositions are not (or, at least, need not be) simply vistigial remains of the simpler dyadic view of belief, playing no particularly important or indispensible role in the triadic theory. Propositions are here identified with the *contents* of beliefs; meanings are identified with *manners* in which one may hold beliefs.

Presumably, *what* is believed is just as important as *how* it is believed. Furthermore, it is, presumably, in terms of propositions (or, at least, mostly in terms of propositions that we evaluate claims concerning the retention of belief and claims that two people have the same belief.

As I have characterized it, the triadic theory is a (metaphysical) theory about the nature of belief; as such, it is not directly concerned with the semantical problem of truth conditions for ascriptions of belief. Some partisans of the triadic view[9] have suggested that even though belief is to be understood as above, an ascription ⌜a believes that S⌝, a a term, S a sentence, is true exactly if a's referent believes (*viz.*, has as the content of a belief) the proposition expressed by S. Nonetheless, if we suppose the triadic view of belief to be correct, it is natural to think that the triadic nature of belief might be reflected in ascriptions of belief. One way in which such a reflection might occur is this: Some ascriptions of belief may imply, not only that a particular proposition is an object of belief, but that it is believed in a certain way – that is, that it is believed under a meaning of a certain sort.

It is the task of the remainder of this paper to make it plausible that this indeed *is* true of ascriptions of belief. Of course, the account of belief ascriptions to be developed will be consistent with the thesis of direct reference, as we have characterized it. We will begin by developing, in the next section, a treatment of so-called *de se* ascriptions of belief.

II

A *de se* ascription of belief is one of the form of

(1) a believes himself to be F.

where a is a term, and *is* F is a predicative expression.[10] It is widely acknowledged that a *de se* ascription of the form of (1) is not implied by the corresponding *de re* ascription, that of the form of

(2) There is an x such that x is identical with a and x believes that x is F.

I presume that the reader is acquainted with the standard arguments that sentences of the form of (2) don't imply ones of the form of (1);[11] I will assume here that such an implication does indeed fail. It is, initially, difficult to see how an advocate of the thesis of direct reference could deny that this implication fails.

The problem is this: Consider a particular *de se* ascription – say, 'John believes himself to be wise' – and the corresponding *de re* ascription, 'There is an x such that x is John and x believes that x is wise'. Under what conditions is the *de se* ascription true? Presumably, it is used to ascribe to John belief in the proposition John would express by saying 'I am wise'; thus, one thinks, it will be true iff John believes that proposition. But on the thesis of direct reference, this is the proposition that is expressed by a sentence of the form $\ulcorner d$ is wise$\urcorner$, d a directly referential term denoting John. It seems that the advocate of direct reference ought to hold that John's believing such a proposition is both necessary and sufficient for the truth of the corresponding *de re* ascription. After all, '$(\exists x)$ (x = John and x believes that x is wise)' seems to ascribe to John a belief in what 'x is wise' expresses, when John is assigned to 'x'; but the free variable, under an assignment, seems to be the paradigm of a directly referential term.

A straightforward account of why the implication does not go through is motivated by the triadic theory of belief. For one can say that the ascription 'John believes himself to be wise' implies that John believes the proposition that he would express by saying 'I am wise' in a certain way – roughly, under the meaning of 'I am wise'. Pretty obviously, given the triadic theory of belief, John can believe this proposition under meanings other than that of 'I am wise': Perhaps, for example, John sees a reflection of himself, doesn't know that he sees himself, accepts the meaning of 'He [John demonstrates his reflection] is wise', but doesn't accept the meaning of 'I am wise'. As long as John believes the proposition under some meaning, the *de re* ascription is true. Thus, we have an explanation of how it is that the *de re* ascription fails to imply the *de se* ascription.

A general treatment of *de se* ascriptions may be developed along the following lines. First, let us introduce some structure to meanings. Instead of thinking of a meaning as simply a function from contexts to propositions, think of it as a pair $\langle\langle s_1, \ldots, s_n\rangle, M^n\rangle$ ($n \geqslant 0$), where each s_i is a (demonstrative) term-meaning – a function from contexts to individuals – and M^n is an n-place predicate-meaning – a function from contexts to n-place properties. (I will, for the sake of expediency, identify n-place properties with functions from n-tuples of possible individuals to sets of possible worlds; propositions with zero-place properties – *viz.*, sets of worlds.) The proposition such a meaning yields in a context c is, of course, the proposition p such that w is in p exactly if w is in $[M^n(c)](\langle s_1(c), s_2(c), \ldots, s_n(c)\rangle)$.

Note, now, that we can 'partially interpret' such meanings, relative to a context. For example, if we start with a meaning $m = \langle\langle s_1, s_2\rangle, M^2\rangle$ and a context c, we can 'plug in' the values of s_1 and M^2 in c to get a 'reduced meaning' $m' = \langle\langle s_2\rangle, P^1\rangle$, P^1 the one-place property such that $w \in P^1(u)$ iff $w \in [M^2(c)](\langle s_1(c), u\rangle)$. The reduced meaning m', in turn, corresponds to the function from contexts to propositions which applied to a context c' yields the proposition that the value of s_2 in c' has P^1.

The basic intuition behind the general treatment of *de se* ascriptions we propose is this: A *de se* ascription

(3) *a* believes himself to be *F*

is true exactly if *a*'s referent believes the proposition that he is *F* (*viz.*, the proposition that he has the property which is expressed by ⌜is *F*⌝ relative to the context at which we interpret (3)) under a meaning *m* which has as one of its reduced meanings $\langle\langle\{I\}\rangle, F\rangle$, where $\{I\}$ is the meaning of '*I*'. This, in turn, will be true precisely if *a*'s referent accepts a meaning which is the meaning of a sentence of the form ⌜$\phi(I)$⌝, where $\phi(x)$ expresses, relative to his context, the property *F*. When someone believes a proposition under such a meaning, we will say that he self-attributes the property, allowing us to state our view in summary form as: (3) is true exactly if *a*'s referent self-attributes the property expressed by ⌜is *F*⌝.[12]

The semantical details of a formalization of our approach to *de se* ascriptions are not particularly complex. Syntactical details, however, are slightly subtle.

Consider, to begin with, the behaviour of 'believes' in *de dicto* and *de re* ascriptions and in *de se* ascriptions. In the first two sorts of ascriptions, the belief operator – use 'B^r' to represent it – appears to operate on an n-place predicate ($n \geqslant 0$) to yield an $n + 1$-place predicate. For example, 'at the level of logical form', 'B^r' combines with 'x loves y' to yield 'zB^r(x loves y)'. The belief operator in *de se* ascriptions, on the other hand – let us use 'B^s' to represent it – apparently combines with an n-place predicate ($n > 0$) and a specification of an argument place to yield an n-place predicate. Thus, for example, applying 'B^s' to 'x loves y' and specifying the first argument place seems to yield something along the lines of 'zB^s (he himself loves y)'.

Of course, given that we do not want *de se* ascriptions to be implied by the corresponding *de re* ascriptions, we cannot assume that something like 'zB^s (he himself loves y)' is reducible to an expression involving 'B^r' and other syntactic operations. For example, we would not want to identify 'zB^s(he himself loves y)' with the result of applying the operation 'identifying the first two argument places' to 'zB^r(x loves y)'. For the latter object – 'zB^r(z loves y)' – will be true, relative to an assignment f, precisely if $f(z)$ believes *de re*, with respect to $f(z)$ and $f(y)$, that the former loves the latter.

Thus, we will use two distinct belief operators, 'B^r' and 'B^s', in our formalization. 'B^r' will, as is usual, take a sentential complement. We will, however, have 'B^s' take as complement a 'property abstract' (something of the form $\ulcorner\hat{x}(\phi)\urcorner$, ϕ a sentence). The reasons for treating 'B^s' in this way have, for the most part, to do with elegance in presentation. We could, in principle, allow 'B^s' to take a sentential complement, so long as we introduced apparatus for indicating what argument positions in an embedded sentence are 'specified argument places' in the sense indicated above. Such a treatment, however, is messier than need be.

It should be stressed that the decision to treat the *de se* belief operator in this way does *not* constitute surrender of the view that the objects of belief (*viz*., the contents of belief, in the sense of Section I) are uniformly propositions, nor does it make it at all inappropriate to say that something of the form $\ulcorner\alpha B^s\hat{x}(\phi)\urcorner$ is (a representation of) an ascription of *belief*. Our semantics will take a formula of the form of $\ulcorner\alpha B^s\hat{x}(\phi)\urcorner$ to be true precisely if α's referent believes a *proposition* under a meaning m which has $\langle\langle\{I\}\rangle, \overline{\hat{x}(\phi)}\rangle$ as a reduced meaning, where $\overline{\hat{x}(\phi)}$ is the property the semantics associates with $\hat{x}(\phi)$. Furthermore, as we will show, a *de se* ascription will, in this treatment, imply its corresponding *de re* ascription (and thus imply that a certain proposition is believed), although the converse implication, of course, will not hold.

The vocabulary and formation rules for our treatment are as follows. As primitive vocabulary items we have: A denumerable set $V = \{x_1, x_2, \ldots\}$ of variables; denumerable sets $Y = \{y_1, y_2, \ldots\}$ and $T = \{t_1, t_2, \ldots\}$ of demonstrative terms (used to represent, respectively, uses of second person singular 'you' and third person singular demonstratives such as 'he', 'she', 'that', etc.); the singular term: I; for each n, a denumerable set F^n of n-place predicates; the truth functors: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$; the belief predicates: B^r, B^s; the abstraction operator: $\hat{}$; the quantifiers: $\exists, \forall$; and, as punctuation, '(', ')'. We use D to name the set of demonstratives of the language, the set $Y \cup T \cup \{I\}$; $\mathcal{T}$, the set of terms, is $D \cup V$.

The definition of well-formed formula is:

1. If $\Pi \in F^n$ and $\alpha_1, \ldots, \alpha_n \in \mathcal{T}$, then $\ulcorner \Pi^n \alpha_1 \ldots \alpha_n \urcorner$ is a formula.
2. If ϕ and Ψ are formulas, then $\ulcorner \neg(\phi) \urcorner$, $\ulcorner (\phi) \wedge (\Psi) \urcorner$, $\ulcorner (\phi) \wedge (\Psi) \urcorner$, $\ulcorner (\phi) \rightarrow (\Psi) \urcorner$ and $\ulcorner (\phi) \leftrightarrow (\Psi) \urcorner$ are formulas.
3. If ϕ is a formula, $\alpha \in V$, then $\ulcorner \exists \alpha(\phi) \urcorner$, $\ulcorner \forall \alpha(\phi) \urcorner$ are formulas.
4. If ϕ is a formula, $\alpha \in \mathcal{T}$, then $\ulcorner \alpha B^r(\phi) \urcorner$ is a formula.
5. If $\alpha \in \mathcal{T}$ and Γ is a proper abstract, then $\ulcorner \alpha B^s \Gamma \urcorner$ is a formula, where a proper abstract is any expression of the form $\ulcorner \hat{\alpha}(\phi) \urcorner$, ϕ a formula and α a member of V which occurs freely in ϕ.
6. These are all the formulas.

Before discussing the semantics, it is perhaps worthwhile to discuss the intuitive readings of those expressions which are formulas in virtue of clauses (4) and (5). Consider the following well-formed expressions of the language:

(4) $IB^r(IB^r(FI))$,

(5) $IB^r(IB^s\hat{x}(Fx))$,

(6) $IB^s\hat{x}(xB^r(FI))$,

(7) $IB^s\hat{x}(IB^r(Fx))$,

(8) $IB^s\hat{x}(xB^s\hat{x}(Fx))$.

The semantic differences among these can be brought out as follows. Read 'Fx' as 'x is wise'; imagine me to be standing by a mirror. (4) through (8) can be understood as representing different readings of "I believe that I believe that I am wise", the difference in readings corresponding to different sorts of meanings under which I might hold the belief: (4) will be true simply if I hold a belief under the meaning of 'He believes [*de re*] that he is wise' (occurrences of 'he' always accompanied by a demonstration by me of my reflection); (5) corresponds to belief under the meaning of 'He believes himself to be wise'; (6) under 'I believe that he is wise'; (7), 'He believes that I'm wise'; (8), 'I believe myself to be wise'.[13]

We define an interpretation for the language as a quartet $M = \langle U, W, C, V\rangle$ which obeys the following strictures:[14]

1. U, W, and C are non-empty and disjoint sets (which, intuitively, represent possible individuals, worlds, and contexts, respectively).
2. (a) Associated with each member c of C is four-tuple $\langle c_A, c_W, c_Y, c_T\rangle$,
   (i) $c_A \in U$ (c's agent),
   (ii) $c_W \in W$ (c's world),
   (iii) c_Y and c_T are denumerable sequences of members of U (the potential addressees and demonstrata of c).
   (b) $c = c'$ iff $c_A = c'_A$, $c_W = c'_W$, $c_Y = c'_Y$, and $c_T = c'_T$.
   (c) No world contains distinct contexts with the same agent.
3. V is a function which assigns
   (a) a member of $((\mathscr{P}(W))^{U^n})^C$ to each member of F^n, for each n;
   (b) sets of meanings to each member of C, where a meaning is a pair $\langle\langle s_1, \ldots, s_n\rangle, M^n\rangle$ $(n \geqslant 0)$, each $s_i \in U^C$ and M^n a member of $((\mathscr{P}(W))^{U^n})^C$.

A word on the workings of V is perhaps in order here. V's assignments to predicate letters are, intuitively, predicate-meanings (taken to be functions from contexts to properties). V's assignments to contexts are to be understood as representing the class of meanings under which the agent of the context holds beliefs; in the terminology of Section I, $V(c)$ is the set of meanings which c_A accepts. Note that, for each context c, $V(c)$ determines a set of propositions, a proposition p being in the set so determined by $V(c)$ exactly if, for some m in $V(c)$, m, completely interpreted relative to c, yields p. These, of course, are the propositions which are objects of belief of the agent of c.

We must, in order to give a definition of truth, characterize the conditions under which the agent of a context self-attributes a property. This we do using the notion of a reduced meaning, introduced above. Where $M = \langle\langle s_1, \ldots, s_n\rangle, M^n\rangle$ is a meaning, a reduced meaning corresponding to M, relative to a context c, is any function in $\mathscr{P}(W)^C$ which results (in the way indicated above) by interpreting M^n and one or more of the s_i, relative to c. An i-reduced meaning is any reduced meaning such that (a) not all the s_i's are interpreted; (b) the only s_i's not interpreted are $\{I\}$ ($\{I\}$, of course, is the function which yields c_A, when applied to a context c). Where M is a meaning, we denote the set of i-reduced meanings of M, relative to c, by $M^{i,c}$. A member M_1 of $M^{i,c}$ is said to attribute a one-place property P just in case, for any context c' and world w

$$w \in M_1(c') \text{ iff } w \in P(c'_A).$$

When an $M_1 \in M^{i,c}$ and property P are so related, we write: $P \in [M^{i,c}]$. We can now say that the agent of a context c self-attributes the property P precisely if there is an M in V_c such that $P \in [M^{i,c}]$.

To define truth and denotation in an interpretation (reference to which is continually suppressed), we proceed as follows. The denotation of a term α, relative to a context c, assignment (member of U^V)f, and world w (write: $|\alpha|_{cfw}$) is defined: $f(\alpha)$, if $\alpha \in V$; c_A, if $\alpha = I$; C_{T_i}, if α is t_i; C_{Y_i}, if α is y_i. We begin the definition of ϕ, taken relative to c and f, is true at w (write: $cf[\phi]w$) as follows:

1. $$cf[\Pi^n\alpha_1 \ldots \alpha_n]w \text{ iff } w \in [V(\Pi^n)(c)]\,(\langle|\alpha_1|_{cfw}, \ldots, \alpha_n|_{cfw}\rangle)$$

2. $$cf[(\phi) \wedge (\Psi)]w \text{ iff } cf[\phi]w \text{ and } cf[\Psi]w.$$

And so on, for the other truth functors.

3. $$cf[\exists\alpha(\phi)]w \text{ iff } \exists u(u \in U \text{ and } cf^{\alpha}_{u}[\phi]w).$$

Analogously for $\forall\alpha(\phi)$.

4. $$cf[\alpha B^r(\phi)]w \text{ iff } \exists c'(c'_A = |\alpha|_{cfw}\ \&\ c'_w = w\ \& \\ \exists m(m \in V(c')\ \&\ m(c') = \{w' \mid cf[\phi]w'\})).$$

$m(c')$ here is the proposition yielded by m in c', defined as above.

The intuitive content of clause (4) is this. $\alpha B^r(\phi)$, taken relative to c and f, is true exactly if: There is a meaning m such that α's denotatum accepts it (formally: $m \in V(c')$, c' the context of α's denotatum), and m yields, relative to c', that proposition expressed by ϕ relative to c. Note that this clause has the result (given that a person believes a proposition p if he accepts a meaning which yields p relative to his context) that $\alpha B^r(\phi)$ is true iff what α denotes believes the proposition expressed by ϕ.

Let $\hat{\alpha}(\phi)$ be a proper abstract. We say that P is the implied property of $\hat{\alpha}(\phi)$, taken relative to c and f, if and only if P is the one-place property such that, for all u and w,

$$w \in P(u) \text{ iff } cf^{\alpha}_{u}[\phi]w.$$

We use $\overline{\hat{\alpha}(\phi)}^{cf}$ to denote the implied property of $\hat{\alpha}(\phi)$, taken relative to c and f. We may complete our definition of truth by saying that a *de se* ascription $\alpha B^s\hat{\beta}(\phi)$, taken relative to c and f, is true at w precisely if: α's denotatum believes a proposition under a meaning which has, as one of its i-reduced meanings, one which attributes $\overline{\hat{\beta}(\phi)}^{cf}$ – that is, just in case α's denotatum self-attrubutes $\overline{\hat{\beta}(\phi)}^{cf}$. Formally, we have

5. $cf[\alpha B^s\hat{\beta}(\phi)]w$ iff $\exists c'(c'_A = |\alpha|_{cfw}$ & $c'_w = w$ & $\exists m(m \in V(c')$ & $\overline{\hat{\beta}(\phi)}^{cf} \in [M^{i,c'}])$.

These semantics adequately capture the view of the truth conditions of *de se* ascriptions discussed at the beginning of this section. In particular, they have the consequence that a *de se* ascription implies (what we will presently define as) its corresponding *de re* ascription, although the converse implication does not hold. Thus, something of the form $\alpha B^s\hat{x}(\phi)$ involves an ascription of *belief*: The ascription is true only if α's denotatum believes the proposition ϕ expresses, when the denotatum of α is assigned to x.

We define the *de re* ascription corresponding to a *de se* ascription $\Psi = \alpha B^s\hat{x}(\phi)$ as follows. Let v be the least (i.e., with smallest subscript) variable not occurring in $\alpha B^s\hat{x}(\phi)$. The *de re* ascription corresponding to Ψ is then

$$\exists v(v = \alpha \wedge vB^r(\phi')),$$

where ϕ' is ϕ with all free occurrences of x replaced by v. (We of course understand the expression $\hat{\alpha}$ to bind free occurrences of α within its scope.) Thus, for example, corresponding to

$$IB^s\hat{x}_1(x_1B^s\hat{x}_1(Fx_1))$$

is

$$\exists x_2(x_2 = I \wedge x_2B^r(x_2B^s\hat{x}_1(Fx_1))).$$

It follows fairly directly from the above definitions that whenever a *de se* ascription, taken relative to c and f, is true at w, then so is its corresponding *de re* ascription. Of course, the converse does not hold. For example, if $V(c')$ consists solely of the meaning of 'Ft_5', 't_5' denotes c'_A relative to c',

$$\exists x_1(x_1 = I \wedge x_1B^r(Fx_1))$$

will be true, relative to c' and an assignment f, at c_w, but

$$IB^s\hat{x}(Fx)$$

will not.

There is a sense in which the semantics allows us to dispense with 'B^r' and make do with only 'B^s' as a belief predicate. For we can define 'B^r' using a schema along the lines of

$$\alpha B^r(\phi) = df \qquad \alpha B^s\hat{\beta}(\beta = \beta \wedge \phi)$$

With some minor tinkering, this would be an adequate definition. (The tinkering required is this: As it stands, it's not the case that

$$\alpha B^r(\phi)$$

and

$$\alpha B^s\hat{\beta}(\beta = \beta \wedge \phi)$$

always agree in truth value, since (speaking very loosely) the latter's truth requires that the believer believe under the meaning of $\ulcorner I = I \wedge \phi \urcorner$, while the former requires simply belief under the meaning of ϕ. Now, although these meanings are identical when *conceived as functions* from contexts to propositions, they are not identical when conceived, as in our semantical system, as ordered n-tuples of the meanings of constituent expressions. Thus, to implement the above definition, we'd need to impose a requirement on the function V in our models to the effect that $\{I = I \wedge \phi\} \in V(c)$, if $\{\phi\} \in V(c)$.)

However, such a definition has little, philosophically, to recommend it. The possibility of such a definition does not show that, in our regimentation, belief *de dicto* and *de re* are kinds of, or are reducuble to, belief *de se*. (What it shows, I think, is that our system is committed to the thesis that anyone who believes a proposition p believes that he's himself and p, and the converse.) And it is certainly not the case that such a definition is what authors like Lewis [4] and Chisholm [1] have in mind when they suggest that belief *de re* is a kind of belief *de se*.

To take Lewis as an example: His view is that to believe *de re* of u that she's F is to self-ascribe the property *bearing R to one and only one thing, a thing that's F*, where R is a 'suitable' relation and one indeed bears R to u and u alone. On such a view, *de re* belief isn't to be represented via quantification into the belief context (as we have represented it), nor will someone with such a view be sympathetic with our treatment of belief ascriptions involving demonstratives other than "I" (which is, in part, designed to represent such ascriptions as ascriptions of belief in propositions 'singular' with respect to the referents of the demonstratives). What is critical to regimenting Lewis' view is not eliminating 'B^r' in favor of 'B^s' (although that's involved), but giving a procedure for representing ascriptions, which appear to involve quantifying in, as not involving it.

Thus, we will preserve the operator "B^r", devoting the next section to a discussion of its semantics.

III

Our approach to *de se* ascriptions of belief is consistent with the view that the contents of *de dicto*, *de re*, and *de se* beliefs are all of the same category: They are all propositions. On the view just formalized, a (use of a) *de se* ascription of belief ascribes belief in a proposition. That is, for any such use u, there is a proposition p such that u is true only if whomever belief is ascribed to, by u, believes p. But, on the approach we have suggested, that is not all such an ascription does; it also tells us something about (it implies that there is a particular) way in which belief is held.

Why should this be true only of *de se* ascriptions? Why, indeed: I believe that this is also true of *de re* ascriptions of belief.[15] I will argue in this section that there are pairs of *de re* ascriptions which ascribe to a person belief in the same proposition (given the theory of direct reference), but diverge in truth value. I will then discuss how a generalization of the semantics developed above can help the advocate of direct reference to account for this.

Consider A – a man stipulated to be intelligent, rational, a competent speaker of English, etc. – who both sees a woman, across the street, in a phone booth, and is speaking to a woman through a phone. He does not realize that the woman to whom he is speaking – B, to giver her a name – is the woman he sees. He perceives her to be in some danger – a run-away steamroller, say, is bearing down upon her phone booth. A waves at the woman; he says nothing into the phone.

If A stopped and quizzed himself concerning what he believes, he might well say

(1) I believe that I can inform you of her danger via the telephone.

(It is understood here, and in the sequel, that uses of 'she' are accompanied by demonstrations of the woman across the street; uses of 'you' are addressed to the woman through the telephone.) A would deny the truth of an utterance by himself of

(2) I believe that I can inform her of her danger via the telephone.

The embedded sentences in (1) and (2) differ only with respect to demonstratives co-referential in the context. Hence (since the embedded sentences do not themselves contain any epistemological operators), if we accept the view that demonstratives are directly referential, we must say the embedded sentences express, relative to the context, the same proposition. Thus, (1) and (2), taken relative to the context, ascribe to A belief in the same proposition.

Surely, however, (1) and (2) diverge in truth value here, (1) being true and (2) being false. One can muster convincing evidence for both these claims. To argue for the truth of (1), for example, we may first note that A surely knows what proposition he expresses when uttering its embedded sentence. For he knows the meaning of the sentence, he is perceiving the referents of the demonstratives therein, and may be said to know of each demonstrative that it denotes the thing perceived. (To forestall one sort of objection, suppose B to be speaking into the phone throughout the example.) Furthermore, the embedded sentence in (1) certainly seems to express something that A believes, namely 'I can tell *this* of the danger of *that* via the phone'. Given all of this, and the fact that A would, sincerely and after reflection, attest to the truth of (1), it seems that we ought to allow that (1) is true.

It does not follow, however, that (2) is true; indeed, it would seem that (2) is certainly not true. One argument one could advance in favor of this claim is this: (2) is true only if A believes that there is someone in danger with whom he can converse via the phone. As the case is set up, there's every reason to think that A does not have this belief. Hence, there's every reason to think that (2) isn't true.

Presently, I will discuss what an advocate of the view that demonstratives are directly referential ought to say regarding cases like the one I've just presented. Before doing so, I will digress in order to consider a case which, superficialy, appears similar to one I've just outlined.

Consider again the situation of A and B. If A stopped and quizzed himself concerning what he believes, he might well sincerely utter

(3) I believe that she is in danger.

but not

(4) I believe that you are in danger.

Many people, I think, suppose that here, again, we have a case in which sentences which ascribe belief to A in the same proposition (given that demonstratives are directly referential) clearly diverge in truth value, (3) being true and (4) being false.

It's clear that if we accept the thesis of direct reference, we must say that the embedded sentences in (3) and (4) express, relative to A's context, the same proposition. But the view – that (3) is true in the context and (4) is not – is, I believe, demonstrably false. In order to simplify the statement of the argument which shows that the truth of (4) follows from the truth of (3), allow me to assume that A is the unique man watching B. Then we may argue as follows:

Suppose that (3) is true, relative to A's context. Then B can truly say that the man watching her – A, of course – believes that she is in danger. Thus, if B were to utter

(5) The man watching me believes that I'm in danger.

(even through the telephone) she'd speak truly. But if B's utterance of (5) through the telephone, heard by A, would be true, then A would speak truly, were he to utter, through the phone

(6) The man watching you believes that you are in danger.

Thus, (6) is true, taken relative to A's context. But, of course,

(7) I am the man watching you.

is true, relative to A's context. But (4) is deducible from (6) and (7). Hence,[16] (4) is true, relative to A's context.

Note that a similar argument can't be used to show that from the claim that (1) is true in A's context, it follows that (2) is true. Consider how we might attempt to construct such an argument. We would have to argue that if A can truly utter (1), then B can truly say that the man watching her believes that he can inform her of her danger via the phone. That is, we would have to claim that if A can truly utter (1), then B can truly utter

(8) The man watching me believes that he can inform me of my danger via the phone.

Here, I think, the new argument goes awry. For it follows, from the claims that B can truly utter (8) and that A is the man watching B, that A believes that there is someone in danger who's such that he can tell her of her danger via the phone. But, as the case is set up, this is not so. Hence, there's no reason to think that an utterance of (8) by B would be true. (I will discuss this further below.)

Let us now return to the original case. It is clear what we will say about this case, if we accept the view of belief above labelled the triadic view. We will say that A believes the proposition – that B can be informed of her danger via the phone – under the meaning of the embedded sentence of

(1) I believe that I can inform you of her danger via the telephone.

but not under the meaning of the embedded sentence of

(2) I believe that I can inform her of her danger via the telephone.

This analysis shouldn't be terribly puzzling, even given that A understands both sentences and knows of each, and the proposition it expresses, that the former expresses the latter. For, as A doesn't know that his uses of 'she' and 'you' are co-referential, he can hardly be expected to know that the embedded sentences express the same proposition.

Compare, now, the position of A with that of a person X, who is in the same situation as A, but who knows that the woman he sees is the woman to whom he is speaking. X will hold a belief about B under both the meanings mentioned above. He will also differ from A in the following way: There will be a woman whom X believes to have the property *being such that she can be informed of her danger via the phone*. It seems that we cannot explain this difference between A and X in terms of proposition believed, since both of them believe the proposition that B can be informed of her danger via the phone. In order to explain the difference, we must appeal to *how* A and X hold their beliefs. It would seem that to believe the proposition expressed (relative to a context c) by a sentence in which demonstratives occur is to have a *de re* belief with respect to the objects denoted, in c, by the demonstratives in the sentence. If one has a *de re* belief with respect to an object, then one may be said to *attribute* certain properties to the object. However, it does not follow, from the fact that x and y each believe the proposition p expressed in c by a sentence $S(d)$, d a demonstrative occuring in S and denoting u in c, that every property which x attributes to u, in virtue of his believing p, is one which y attributes to u, in virtue of this belief. For which properties one attributes to an object is determined by the meaning under which one's belief is held: X, for example, who believes the proposition, that he can inform B of her danger via the phone, under the meaning of 'I can inform her of her danger via the phone' will attribute to B the property *Being a thing that can be informed of its danger via the phone*: A, who doesn't believe the proposition under the meaning just mentioned, will not attribute this property to B.

If this much be accepted, we have the basis of an answer to the question: How can

(1) I believe that I can inform you of her danger via the telephone.

and

(2) I believe that I can inform her of her danger via the telephone.

diverge in truth value in a context in which their embedded sentences express the same proposition? For we may say: An ascription of belief ⌜a believes that S⌝, S a sentence in which demonstratives occur, not only implies that the proposition expressed by S is believed, but that certain properties are attributed to the referents of the demonstratives in S. What properties the ascription implies are attributed depends, in turn, upon the meaning of S. In the case in question, ascription (2) implies that a property (that assocaited with a use, in this context, of 'I can inform x of x's danger by phone') is attributed which (1) does not imply is attributed. Hence, (1) may true be while (2) is not.

Let us consider how we might give a systematic development of this proposal. In order to simplify matters, we will do this for a language with only a *de re* belief operator; it will be obvious how the treatment would be generalized to a language including a *de se* operator such as that discussed in Section II.

We assume, then, that our language has the same primitive vocabulary as the language of Section II, minus the B^s operator and the abstraction operator; the formation rules are identical to those of Section II, save the omission of the clause of the *de se* operator. We preserve the definitions of interpretation, denotation, and the clauses of the truth definition for atomic, truth functional, and quantified sentences. We now need to characterize, in terms of the formal structure, two things: When an individual, in believing a proposition under a meaning, attributes a property, and when a belief ascription, taken relative to a context, implies the attribution of a property.

Let $m = \langle\langle s_1, \ldots, s_n\rangle, M^n\rangle$ be a meaning. The intuitive answer to the question – When does the agent of a context c attribute a property P, in virtue of believing under m? – is as follows. Consider, first of all, what one 'gets' if one (a) replaces M^n with $M^n(c)$ (*viz.*, replaces the meaning M^n with the property which is its value in c); (b) replaces each s_i either with its value in c or with a variable; (c) doesn't replace distinct s_i's with the same variable. Call such entities the *proto-properties* associated with m in c.

(For example, proto-properties associated with

$$m_1 = \langle\langle\{t_1\}, \{y_1\}\rangle, \{F_1^2\}\rangle$$

– which could be identified with the meaning of "$F_1^2 t_1 y_1$" – in a context in which "t_1" denotes u, "y_1" denotes u' and "F_1^2" denotes P are

(i) $\langle\langle u, x\rangle, P\rangle$,

(ii) $\langle\langle x, u'\rangle, P\rangle$,

(iii) $\langle\langle x, x'\rangle, P\rangle$.

Proto-properties associated with

$$M_2 = \langle\langle\{t_1\}, \{t_2\}\rangle, \{F_1^2\}\rangle$$

in such a context are all of the above and

(iv) $\langle\langle x, x\rangle, P\rangle$.)

To each proto-property there corresponds, in a rather obvious way, a property. For example: to (ii) corresponds the one-place property P^1 such that $w \in P^1(u_1)$ iff $w \in P(\langle u_1, u'\rangle)$; to (iii) corresponds the two-place property P^2 such that $w \in P^2(\langle u_1, u_2\rangle)$ iff $w \in P(\langle u_1, u_2\rangle)$; to (iv) corresponds the one-place property P^3 such that $w \in P^3(u_1)$ iff $w \in P(\langle u_1, u_1\rangle)$.

We can now answer our initial question thus: An agent attributes a property P, in virtue of holding a belief under a meaning m iff P corresponds to one of the proto-properties associated with m relative to the agent's context. We will write

P $\quad\quad P \in P(m, c)$

for: the agent of c attributes P, in vritue of holding a belief under m.

A fully rigorous characterization of the above notion would disperse with the notion of a variable in the construction of proto-properties. It is easy enough to give such a characterization; we henceforth assume that the predicate $P(m, c)$ has been so defined in terms of our model structure. We now need a way to get from a sentence (taken relative to a context and an assignment) used to ascribe belief to the set of properties it implies the believer attributes. One way of doing this is as follows. Consider a sentence ϕ; let $\alpha_1, \ldots, \alpha_n$ be a complete enumeration of those demonstratives and variables (which occur freely) in ϕ. Let $v_1, \ldots, v_n$ be variables which do not occur in ϕ. We say that Ψ is a frame of ϕ just in case ϕ is the result of replacing one or more of the α_i's with v_i's, subject to the restriction that distinct α_i's are replaced with distinct v_i's.

Thus, for example, consider the sentences

(i) $\qquad F_2^2 t_1 y_1,$

(ii) $\qquad F_2^2 t_1 t_1.$

Frames of (i) are: $F_2^2 t_1 x_1, F_2^2 x_1 y_1, F_2^2 x_1 x_2$; frames of (ii) are the above and $F_2^2 x_1 x_1$. Note that this last is not a frame of (i).

We say that a sentence ϕ implies the attribution of the property P^n, relative to c and f, just in case there is a frame ψ of ϕ, obtained by substituting the n distinct variables $v_1, \ldots, v_n$ for terms in ϕ and, for every w and $u_1, u_2, \ldots, u_n$:

$$cf^{u_1 u_2 \ldots u_n}_{v_1 v_2 \ldots v_n}[\Psi]_w \quad \text{iff} \quad w \in P^n(\langle u_1, u_2, \ldots, u_n \rangle).$$

We define the attribution class of a sentence ϕ, relative to c and f, as the set of those properties such that ϕ implies their attribution, relative to c and f; we denote this class with $A(\phi, c, f)$.

We now define truth for *de re* ascriptions of belief:

$$\begin{aligned} &cf[\alpha B^r \phi]w \text{ iff } \exists c'(|\alpha|_{cfw} = c'_A \,\&\, c'_w = w \,\&\, \exists m(m \in V'_c \,\&\\ &m(c') = \{w' | cf[\phi]w'\} \,\&\, (f)(f \in A(\phi, c, f) \rightarrow f \in P(m, c')))), \end{aligned}$$

where $m(c')$ is the proposition expressed by m relative to c'. Verbally, these truth conditions amount to this: $\alpha B^r \phi$, relative to c and f, is true exactly if there is a meaning m such that (i) $|\alpha|_{cfw}$ believes a proposition under m; (ii) m yields, relative to $|\alpha|_{cfw}$'s context, whatever ϕ expresses, relative to c and f, and (iii) whatever properties ϕ implies are attributed are such that belief under m requires their attribution.

It is easy to show that, given this semantics, representatives of sentences (1) and (2) can diverge in truth value relative to a context in which their embedded sentences express the same proposition.[17] On the other hand, the semantics validates, the claim, for which we argued above, that in any context in which the uses of 'she' and 'you' in

(3) I believe that she is in danger.

and

(4) I believe that you are in danger.

are co-referential, the truth of (4) is implied by the truth of (3).

It is, perhaps, worth discussing sentences (3) and (4) again. Many people, even after a rehearsal of the argument given above – that (4) is implied by (3) – are still uncomfortable with the claim that both (3) and (4) are true. A virtue of the semantics just presented, I think, is that it can be used to motivate an explanation of why the intuition, that (3) and (4) diverge in truth value, is so persistent.

Take a finite set of sentences and conjoin them; form what we called a frame of the result. (For example, if you start with {that$_2$ is sad, you$_3$ will make that$_4$ happy if that$_2$ helps you$_3$}, you will end up with something along the lines of 'x_2 is sad $\wedge$ x_3 will make x_4 happy if x_2 helps x_3'.) Call the property associated with such a sentence a *picture*; if all the members of the initial set are sentences, the meanings of which are accepted by an agent u, say that the resulting property is a picture *held by* u.[18]

The intuition motivating our semantical account is that an ascription is true provided is ascribes belief in a proposition which is believed and the ascription doesn't imply anything false about what pictures are held by the believer. Since sentence (4), as used by A, does not *when taken by itself* imply anything false about what pictures A holds, (4) so taken is true, since A believes B to be in danger.

Note, now, that a set of belief ascriptions may (conventionally) imply things about the pictures a believer holds that the conjunction of the members of the set does not (strictly) imply.[19] For example, the use of the ascription 'A believes that you$_1$ are unhappy because she$_2$ spurned you$_1$' in a context in which the ascription 'A believes that she$_2$ loves a Greek' has been used (and no one has disputed the truth of the latter ascription) will imply that A holds the picture associated with 'y loves a Greek and x is unhappy because y spurned x'. Both ascriptions can be true, even if A doesn't hold this picture; however, their joint use, in such a case, would be very misleading.

In general, we tend to avoid using an ascription $\ulcorner\alpha$ believes that $\phi\urcorner$, if an ascription $\ulcorner\alpha$ believes that $\psi\urcorner$ is assumed by all the parties to the conversation to be true (and we know this), and we think that the person to whom belief is being ascribed does not hold pictures associated with frames of $\ulcorner\phi$ and $\psi\urcorner$. Likewise, we will find an ascription $\ulcorner\alpha$ believes that $\phi\urcorner$ bizarre or objectionable if it is assumed by those conversing that the ascription $\ulcorner\alpha$ believes that $\psi\urcorner$ is true and we have good reason to think that the believer doesn't hold all the pictures associated with $\ulcorner\phi$ and $\psi\urcorner$.

All of this, I beleive, helps to explain why some find the assertion, that A's use of

(4) I believe that you are in danger.

is true, counter-intuitive, even after a rehearsal of the argument that A's use of (4) cannot be false if his use of (3) is not. For as we have just seen, without qualification and explanation, the claim that (4) is true relative to A's context is very misleading. For obviously, in the case under consideration

(9) I believe that I am talking to you.

is true relative to A's context. Thus, without further qualification, the claim that (4) is true implies that

(10) I believe that I am talking to someone who is in danger.

is true, relative to A's context. But, obviously (10) is not thus true.

I close with some observations on the semantical theory suggested in this paper. According to this theory, ascriptions of belief are primarily, but not exclusively, vehicles for making reports about the content, as opposed to the manner, of belief. That ascriptions are primarily used to make reports about content ought not be surprising. For, first of all, we are very often not in a position to say how a belief is held, although we know that it is held. (For example, one may know that Hank believes that Will spies, but not whether he accepts the meaning of 'that spies' or 'Will spies'.) Furthermore, it is often quite irrelevant to our purposes to specify how a proposition is believed.[20] Finally, we often cannot say how belief is held in any perspicuous way, even though we know. (Consider: Hank, Bernie and Sally all believe that I am a spy.)

None the less, ascriptions of belief are, to a limited extent, used to report how belief is held. Indeed, in the semantics for ascriptions of belief suggested in this paper, the belief operator is construed as operating on sentential meanings, and not simply as an operator on the propositions which meanings, relative to a context, have as values. I have focused here upon relatively simple aspects of sentential meaning, in an attempt to make a case for the claim that, by construing the belief operator as an operator on meanings, as opposed to propositions, we can generate plausible solutions to semantical puzzles associated with the (quite plausible, I believe) theory of direct reference. If the approach taken here strikes the reader as not without merit, he or she will, I hope, consider the question of how it is to be given the extensions and refinements it requires in order to yield a fully satisfactory theory.[21]

NOTES

[1] Kaplan [2], p. 1. Henceforth, I will use 'demonstratives' as shorthand for 'demonstratives and indexicals'.

[2] Thus, for example, Kaplan, in the section of [2] entitled "Adding 'Says'" suggests truth conditions for (indirect discourse) ascriptions of belief which have the effect of making something of the form $\ulcorner\alpha$ believes that $\phi\urcorner$ true exactly if α's referent believes (under any meaning whatsoever) the proposition the semantics assigns to ϕ. Elsewhere in [2], Kaplan claims that all (non-quotational) operators of English are 'at most intensional' – *viz.*, they all can be construed as operating on (the formal representatives of) propositions.

[3] An excellent discussion of what the thesis of direct reference does and does not imply can be found in Salmon [8].

[4] I am adopting here some of the terminology and semantic assumptions of Kaplan's [2] and [3].

[5] I have argued in [7] that tense operators can be given an adequate semantical treatment only on the assumption that they operate on the meanings of, not simply on the propositions expressed by, sentences. Thus, questions about 'believes that' to one side, I think the possibility, of there being an operator such as O, is not at all idle.

[6] See [2], [5], and [6].

[7] This is closer to Kaplan's view than Perry's. On Perry's account, the second term in the relation is what Perry calls a *belief state*, which is a mental state individuated (in part, at least) in terms of the sentence types (or meanings thereof) which an agent in that state accepts, where *acceptance* is a technical term with a meaning related to (but probably not identical with) the meaning the term is accorded below in the text.

I characterize the triadic theory as in the text because I find it easier to motivate the formalism of Sections II and III in terms of such a characterization.

[8] I ought to say something here about what these meanings are, and how they differ; what needs to be made clear is what the meaning of terms like 'you' and 'she' is.

I presume the following (and do not suggest that it is an original view; it is a version of Kaplan's own view). There are what we might call 'modes of demonstrating' things and 'modes of addressing' things. These modes are such that the same mode can be used in different contexts or several times in one context. It is only when 'she' is accompanied by a mode of demonstrating ('you' is accompanied by a mode of addressing) that it refers to an object. Furthermore, although 'she' plus mode m of demonstrating ('you' plus mode m' of addressing) may pick out different objects in different contexts, "she" accompanied by one mode of demonstrating picks out the same object every time it is used in a context; analogously for 'you'.

The meaning (in Kaplan's sense of meaning as character) of 'she', then, is roughly this: 'she', accompanied by a mode of demonstrating, functions as a directly referential term; it denotes, relative to a context, what its accompanying mode of demonstrating demonstrates.

Thus, in giving formal representatives for sentences such as those mentioned in the text, what we really represent is the sentence type and aspects of the modes of demonstration or address. (For we wish to be able to assign the representatives of propositions to the formal representatives of sentences; the sentences being represented don't express propositions, on the view assumed here, unless accompanied by modes of demonstration or address.) We thus represent two occurrences of 'she' (of 'you') with the same term if and only if they are accompanied by the same mode of demonstration (or address).

These details will be germane to the view of *de re* belief ascriptions discussed in Section III.

[9] For Kaplan's views, see note 2. Perry has suggested in conversation that he accepts something along the general lines of the semantical view expressed in the sentence to which this is a footnote.

[10] As a referee pointed out, it is misleading to single out (1) as 'the form' of *de se* ascriptions in English. This is, firstly, because sentences of the form ⌜I believe that I am F⌝ seem, at least sometimes, to be used to ascribe *de se* belief and sometimes merely *de re* belief. Secondly, some sentences (e.g., 'I believe that Edwina will build a house near mine') seem to be used to report belief *de se* but are neither of the form of (1) nor such that they have a colloquial equivalent of the form of (1). (A further worry is whether or not (1) has a reading on which it is equivalent to (2); whatever the answer to this question, I do not think it will affect the points made in this section.)

I will persist in speaking as if (1) gave the canonical form of *de se* ascriptions – a fiction which, I hope, is no more harmful in this context than the common fiction, in discussions of belief *de re*, of pretending that ⌜a believes, of b, that she's F⌝ is unambiguously *de re*, while ⌜a believes that b is F⌝ is unambiguously *de dicto*.

[11] A sampler of such arguments is to be found in Chisholm's [1]. It is not my purpose here to defend any particular argument as showing that the implication fails. Rather, I assume that it is *very* plausible that the implication does fail. Given this assumption, the question arises: How could an advocate of the view that demonstratives and indexicals are directly referential account for this?

[12] To those familiar with views of *de se* belief advanced by Chisholm in [1] and Lewis in [4], this will sound somewhat familiar. Chisholm introduces a primitive notion *x directly attributes property P to y* which, according to Chisholm, is necessarily reflexive. Chisholm then says that to believe oneself to be F is to directly attribute F to oneself. Lewis suggests that we understand belief *de se* as the *self-ascription* of property.

There are several important differences between our approach and the approaches of Chisholm and Lewis. We do not hold that properties are the objects of *de se* belief, as do Lewis and Chisholm; we also hold that the objects of all beliefs are of uniform character, unlike Chisholm.

On Chisholm's view, it is somewhat mysterious as to why one can directly attribute properties only to oneself. Indeed, for Chisholm, there is no real correlate of direct attribution, relating distinct individuals and a property: Chisholm's indirect attribution (in terms of which Chisholm defines *de re* belief) is simply a complicated form of direct attribution.

On our view the reflexivity of self-attribution is not mysterious at all: It's reflexive because it involves meanings which contain $\{I\}$. Furthermore, we could define a perfectly analogous notion of indirect attribution, without invoking the notion of self-attribution, if we wished. Indeed, something like this is defined in Section III, below.

We have analogous differences with Lewis, who characterizes belief *de re* in [4] as a kind of belief *de se*. (For Lewis, as for us, the objects of belief are of uniform character; but, unlike us, he takes them to be all properties.)

It is worth noting that the formalization introduced in this section could be used, with some alterations, to regiment Lewis' view. (The major alterations would be to drop the 'B^r' operator introduced below, translating English sentences of the form of *a believes that S*, where *S* involves no reflexivies, as: $aB^s\hat{\alpha}\ (\alpha = \alpha \wedge \phi)$. One would also be required, in a formalization of Lewis' view, to prohibit quantification into 'B^s', and to come up with a scheme to represent *de re* ascriptions. This is discussed at the end of Section II.) This should not hide the fact that there are fundamental differences in motivation between Lewis and ourselves. Beyond those mentioned above, we note that this essay and its formalism is intended to function in the defense of the thesis of direct reference, a thesis which – insofar as it is bound up with what Lewis and Kaplan call 'haecceitism' – is anathema to Lewis.

[13] It is difficult to come up with natural sounding English sentences which unambigously capture these readings. I believe that anyone who takes the notion of *de se* belief seriously will agree that the beliefs represented by (4) through (8) are different beliefs; if the beliefs *are* different, an adequate treatment of belief ascriptions *de se* and *de re* ought to be able to differentiate them, syntactically and semantically.

[14] The semantics presented here is modeled upon that of Kaplan's Logic of Demonstratives; see [2] and [3] for a detailed exposition.

[15] As will become clear below, I consider any ascription of the form ⌜*a* believes that ϕ⌝, which is such that ϕ has explicit occurrences of demonstratives, a *de re* ascription of belief.

[16] I assume a definition of validity such as that which Kaplan gives for his Logic of Demonstratives. (See [3].) I also assume (what is true in that logic) that if *A* follows from *B* and *B* is true in context *c*, then *A* is true in *c*.

[17] We can also show that the semantics validates certain forms of 'quantifying in'. Precisely, given our semantics, we have:

> If β is a member of D which occurs in ϕ, then if $cf[\alpha B^r(\phi)]w$, then $cf[\exists v(\alpha B^r(\phi[\beta/v]))w$, provided that β is free for v in ϕ.

(If our semantics had allowed for the possibility that members of *D* failed to denote in some contexts, this rule would have to be weakened. For simplicity's sake, we have not allowed for this possibility.) That such a rule is sound justifies, in part, the claim that something of the form of $\ulcorner \alpha B^r(\phi) \urcorner$ is a *de re* ascription, provided that ϕ contains a member of D.

Note that not very 'way of quantifying in' is permitted by our semantics. In particular, from

(i) $t_1 = t_2 \wedge IB^r(F^2 t_1 t_2)$

the formula

(ii) $\exists x_1 \exists x_2 (x_1 = x_2 \wedge IB^r(F^2 x_1 x_2)$

follows, but

(iii) $\exists x_1 (x_1 = x_1 \wedge IB^r(F^2 x_1 x_1))$

does *not* follow. Given our reasons for adopting the treatment we have adopted, of course, one would not want (iii) to follow from (i).

[18] Strictly speaking, of course, we can associate properties with open sentences possibly containing demonstratives only relative to a context. My ignoring that here does not effect the point.

[19] I must stress that 'implies' is being used in two senses in this sentence. The first use of 'implies' is quite weak (certainly not the sort of implication which preserves truth). Roughly, the use I intend here is the sort present in (typical) uses of 'His saying that the movie was boring implies that he did not like it'.

[20] Note, however, that it *is* very often important to us to get across that belief is held under a meaning involving $\{I\}$. One reason for this is that we seem to presuppose the truth of a psychological theory which predicts how people will behave when they so believe (and when they have certain desires, etc.). To effectively make use of such a theory in everyday affairs – in particular, to justify predictions of behavior *via* the theory – we need a way to say that a person believes in the relevant way. It is for reasons such as this that English has a *de se* belief operator like that discussed in Section II. That we have no very general need, as we do for beliefs held under meanings involving $\{I\}$, to say that someone holds a belief under the meaning of a sentence involving {that} or {you} explains, I think, the absence of belief operators in English which single out beliefs held under such meanings.

[21] I am indebted to David Auerbach, Edmund Gettier, Richard Grandy, and Harold Levin for comments on my syntax and these semantics. An anonymous referee for this journal also made useful comments on an earlier draft, for which I thank him or her. Part of the work on this paper was done while I held NEH grand FX-28919; I am grateful for this support.

REFERENCES

[1] Chisholm, Roderick, *The First Person*, University of Minnesota Press, 1981.

[2] Kaplan, David, *Demonstratives, Draft #2*, Department of Philosophy, UCLA, 1977.

[3] Kaplan, David, 'On the Logic of Demonstratives', in *Contemporary Perspectives in the Philosophy of Language*, edited by Peter French *et al.*, University of Minnesota Press, 1979.

[4] Lewis, David, 'Attitudes *De Dicto* and *De Se*', *Philosophical Review* 87, 1979.

[5] Perry, John, 'The Problem of the Essential Indexical', *Nous* **13**, 1979.
[6] Perry, John, 'Belief and Acceptance', in *Midwestern Studies in Philosophy, Vol. V*, edited by Peter French *et al.*, University of Minnesota Press, 1980.
[7] Richard, Mark, 'Tense, Propositions, and Meanings', *Philosophical Studies* **41**, 337–351, 1982.
[8] Salmon, Nathan, *References and Essence*, Princeton University Press, 1981.

Department of Philosphy,
DK-50,
University of Washington at Seattle,
Seattle,
Washington,
U.S.A.