Individuating Beliefs

Michael McKinsey

Philosophical Perspectives, Vol. 8, Logic and Language (1994), 303-330.

INDIVIDUATING BELIEFS

Michael McKinsey
Wayne State University

The important work on reference done in the 1960s and 1970s by such philosophers as Donnellan (1970, 1974), Kaplan (1968, 1979), Kripke (1972) and Putnam (1970, 1975), seems to have led to even more astounding conclusions in the 1970s and 1980s, when the earlier semantic results were applied to the interpretation of the cognitive attitudes by such philosophers as McDowell (1977), Evans (1982), Burge (1979, 1982) and Salmon (1986). The earlier work supported the "direct reference theory" of proper names, indexical and demonstrative pronouns, and natural kind terms. According to this theory, sentences containing such terms have "wide" content, that is, express propositions that essentially involve ordinary contingent objects and substances. When this semantic consequence is applied to the interpretation of cognitive attitude predicates of the form 'believes that S', where the sentence S has wide content, we seem to get the further result that the beliefs ascribed by such predicates must themselves also have wide content. From this, many philosophers have gone on to infer that some beliefs, thoughts, and other cognitive acts and states, which would traditionally have been thought to be logically independent of the external world, are in fact individuated by their relations to contingent external objects and substances.

In this paper, I wish to argue that this externalist conclusion about the cognitive attitudes does not follow from the direct reference theory. For to get this conclusion, you have to assume that every *de dicto* cognitive attitude predicate expresses a property that individuates the cognitive state or act ascribed, and this assumption is arguably false. This assumption, which I will call the "Belief-Predicate Assumption," is shared in common by both the defenders of classical forms of internalism and their externalist opponents.[1] The Belief-Predicate Assumption in turn follows from the most firmly entrenched theory of the cognitive attitudes, which I will call the "Proposition Theory." According to

the Proposition Theory, all cognitive attitude acts and states are individuated by their propositional contents, and the cognitive attitude verbs express relations that hold between persons and propositions.

I have described elsewhere a form of cognitive attitude ascription that I call "mental anaphora," and I have argued that this sort of construction provides a counterexample to the Proposition Theory (McKinsey, 1986). I will try to show here that the same phenomenon also provides the basis of an argument against the Belief-Predicate Assumption, and suggests a new general view of the cognitive attitudes and their ascription that explains why the assumption is false.

This general view, which I will describe below, has some important consequences for the philosophy of mind. First, by denying both the Proposition Theory and the Belief-Predicate Assumption, the view can consistently endorse both the direct reference theory and the internalist idea that cognitive acts and states are individuated by their narrow contents. Second, in contrast to other recent defenses of internalism, the view I will describe below achieves this consistency without requiring any kind of *revision*, either of our ordinary concepts of the cognitive attitudes (in contrast to Fodor (1987)) or of our ordinary concept of content (in contrast to White (1982) and Loar (1985)). On the form of internalism I will defend, the distinction between wide and narrow content exists in the meanings of ordinary *de dicto* cognitive attitude predicates. Some of these predicates ascribe wide content and some ascribe narrow content, depending on the semantic properties of the sentence contained in the scope of the attitude operator; but only the predicates that ascribe narrow content express properties that individuate the cognitive acts and states ascribed.

1. Some principles about belief and belief-predicates.

It will be useful to begin by laying out some of the relevant concepts and principles involved in the debate between the internalists and the externalists. For convenience of expression, I will for the most part restrict my discussion to the cognitive state of belief. But I intend the discussion to apply equally well to all "neutral" cognitive acts and states, that is, all cognitive acts and states which, like beliefs, thoughts, intentions, and desires, but unlike knowledge, can have false propositional contents.

By a "belief" I mean a particular concrete psychological state of a person, a state that is numerically distinct from any other person's psychological state, even when the two persons, as we say, "have the same belief." Thus it is slightly misleading to speak of a belief's being "individuated" by the psychological property expressed by a certain belief predicate. For a belief, being a concrete state, may have a nature that outruns characterization by all the psychological properties that are available in our conceptual repertoire. In particular, a concrete belief state may have a *physical* nature that is completely

and essentially characterizable only in terms of certain physical properties. At least, I would prefer not to rule this possibility out in my discussion. Perhaps, we could say, a belief-property individuates a belief, not as a particular concrete object, but rather individuates it *as a belief*. But what could this mean?

Let us say that a given property F is a *belief property* if and only if: it is possible for someone to have F, and necessary that, for any person x, if x has F, then there exists a concrete belief state b such that b is a state of x's having F, and b makes it true that x has F.[2] Then, let us say that a property F *individuates with respect to belief* if and only if: (1) F is a belief property, and (2) for any possible worlds w and w' and any persons x and y, if x has F in w, then y has a belief in w' which is the same belief as the state of x's having F in w if and only if y has F in w'.[3] Of course here, when I use the expression "the same belief as," I don't mean that the beliefs in question are numerically identical, which would be impossible when x≠y. Rather, I mean that the beliefs are "the same" in the sense we have in mind when we say for instance that two persons who both believe that the earth is round thereby "have the same belief."

We can now state the assumption about the meaning of belief predicates that has almost inevitably been made in the literature on the cognitive attitudes, and that leads directly to a conflict between internalism and direct reference. This is

> *The Belief-Predicate Assumption*: For every *de dicto* belief-predicate Φ, if Φ expresses a property F, and it is possible for someone to have F, then F individuates with respect to belief.

By calling a predicate of a language a "*de dicto* belief predicate," I mean that the predicate has the same meaning as a predicate in English of the form 'believes that S', where the sentence S is entirely in the scope of the operator 'believes that'. Often, though perhaps not inevitably, this will mean that substitution of coextensive expressions and existential generalization are invalid modes of inference with respect to the expressions contained in the predicate. Such predicates, that is, form "opaque contexts."

Before going on, we should note the specific historical context in which the Belief-Predicate Assumption has usually been made. The most firmly entrenched view has been that a belief is individuated by *the proposition believed*. We might call this

> *The Propositional Theory of Individuation*: For any possible worlds w and w', persons x and y, and proposition p, if x believes that p in w, then y has a belief in w' which is the same belief as the state of x's believing that p in w if and only if y believes that p in w'.

The Propositional Theory of Individuation is standardly accompanied by the traditional account of the meaning of *de dicto* belief-predicates:

> *The Relation Theory*: For every *de dicto* predicate Φ of the form 'believes

that S' such that Φ expresses a property, there is a proposition p
expressed by the sentence S, and Φ expresses the relational property of
believing that p.

These two theories are usually held in tandem (and in fact are seldom
distinguished from each other). I will follow suit and treat their conjunction as a
single theory, which I will call "the Proposition Theory." Frege (1892) and
Russell after him (1912) made the Proposition Theory the commonly accepted
view, and so it remains to this day.

It is clear that, as I have stated it, the Proposition Theory logically implies
the Belief-Predicate Assumption, which says that every property expressed by a
(consistent) belief predicate is a property that individuates with respect to belief.
But since the converse does not hold, one can consistently accept the Belief-
Predicate Assumption while denying the Proposition Theory.

2. Classical Internalism and Direct Reference.

The issues dividing internalism and externalism are brought into sharpest
focus by considering belief predicates that are constructed by applying the
operator 'believes that' to simple sentences containing ordinary proper names and
indexical pronouns. In the following discussion, I propose to simply adopt the
semantic thesis of direct reference, according to which such sentences express
singular propositions. The thesis of direct reference says that every ordinary
proper name and indexical is what I shall call a "genuine term": a term whose
sole semantic contribution to the proposition expressed by a sentence containing
the term is simply the term's *referent*.

The thesis of direct reference has, I think, been given overwhelming support
by Kripke's (1972) and Kaplan's (1979) work on the modal properties of
sentences containing proper names and indexicals. But this thesis seems to
provide an almost immediate inference to the falsehood of internalism about
belief. Consider any sentence containing an ordinary proper name, such as
'Bismarck was an astute diplomatist'. Since this sentence expresses a
proposition that is singular with respect to the man Bismarck, the predicate
'believes that Bismarck was an astute diplomatist' apparently must express a
property that is relational with respect to the same man. Thus, it appears, any
belief ascribed by this predicate must be individuated by a "wide" property, since
possession of the property expressed by the predicate logically implies the
existence of Bismarck.

It was in all likelihood the plausibility of this inference that provided one of
the main motivations behind both Frege's (1892) and Russell's (1912) *rejection*
of the idea that ordinary names are genuine terms. Both philosophers held what
we may call

The Narrow-Belief Thesis: Any property that individuates with respect to belief is a narrow psychological property.

And both philosophers also clearly made the Belief-Predicate Assumption, according to which every (consistent) *de dicto* belief predicate expresses a property that individuates with respect to belief (the property for both Frege and Russell always being the relational property of believing a certain proposition). These two theses together logically imply what we might call

The Narrow-Predicate Thesis: For every *de dicto* belief predicate Φ, if Φ expresses a property F, and it is possible for someone to have F, then F is a narrow psychological property.

And it is this Narrow-Predicate Thesis that is apparently inconsistent with the direct reference view of ordinary names. For again, if 'Bismarck' is a genuine term, then it certainly seems to follow that the predicate 'believes that Bismarck was an astute diplomatist' expresses a *wide* psychological property.

Simplifying somewhat, let us treat the Frege-Russell view of belief, which I will call "Classical Internalism," as the conjunction of three principles:

Classical Internalism:
   (a) Accepts the Narrow-Belief Thesis.
   (b) Accepts the Proposition Theory; hence,
   (c) Accepts the Narrow-Predicate Thesis.

(Again, the Narrow-Predicate Thesis is really a subthesis that logically follows from the other two *via* the Belief-Predicate Assumption, which in turn is a subsidiary step that follows from the Proposition Theory.)[4]

On the basis of a semantic argument like the one given above from the thesis of direct reference, the externalist rejects the Narrow-Predicate Thesis. Then, making the Belief-Predicate Assumption (presumably on the basis of the Proposition Theory), the externalist infers that the Narrow-Belief Thesis is false and so maintains that some wide psychological properties individuate with respect to belief. Thus:

Externalism:
   (a) Denies the Narrow-Predicate Thesis.
   (b) Assumes the Proposition Theory; and hence
   (c) Denies the Narrow-Belief Thesis.

The classical internalist moves from the Narrow-Belief Thesis to the Narrow-Predicate Thesis *via* the Proposition Theory. Similarly, the externalist moves from the *denial* of the Narrow-Predicate thesis to the *denial* of the Narrow-Belief Thesis, presumably also *via* the Proposition Theory. Proponents of neither view consider the possibility of holding onto Narrow Belief while rejecting both the Proposition Theory and the Narrow-Predicate Thesis. This is the opening spied

by Loar (1985) and other internalists who, like the externalists, are also convinced that the Narrow-Predicate Thesis is false. We might call this view

> *Modified Internalism*:
> (a) Denies the Narrow-Predicate Thesis.
> (b) Asserts the Narrow-Belief Thesis; and hence,
> (c) Denies the Proposition Theory.

Of course, there is more than one form of Modified Internalism: there is Loar's Revisionist form, which adds to its denial of the Narrow-Predicate Thesis the very strong claim that *no* belief-predicate expresses a narrow property; and there is my Non-revisionist form, which merely adds the weak claim that some belief-predicates express narrow properties, and some don't.

3. Why Narrow-Belief?

As I have portrayed the issue so far, the main bone of contention between the internalists and externalists is the Narrow-Belief Thesis, which says that all belief-individuating properties are narrow. But what is really at stake here? Why have so many philosophers who think about the cognitive attitudes found it so natural to follow Frege and Russell in their assumption that our beliefs and thoughts could not be essentially related to objects in the external world, and why should the currently fashionable claim to the contrary seem both so revolutionary and so *prima facie* implausible?

I believe that the main motivation behind the Narrow-Belief Thesis is *epistemological*.[5] Since Descartes, it has seemed undeniable to most philosophers that each of us has a privileged way of knowing about his or her own mental states. It just seems to be a fundamental fact about cognition that whenever we have a thought, belief, intention, or desire, we can in principle come to know *what* we think, believe, intend, or desire just by internal examination, without engaging in an empirical investigation of the external world. Why this should be a fact about the mental, and what this fact tells us about the nature of cognition, are questions that challenge reflection. But that it is a fact seems to me undeniable.

Let us call "a priori knowledge" all knowledge that is obtained without empirical investigation of the external world, including both knowledge of necessary conceptual truths as well as knowledge of contingent facts about one's own mental states.[6] In each case of thought, belief, intention or desire, what we have privileged access to would seem to be the properties that *individuate* those thoughts, beliefs, intentions, and desires. With respect to belief, then, we can express the traditional Principle of Privileged Access as follows:

> *Privileged Access*: Necessarily, if a person has a property φ that individuates with respect to belief, then that person can come to know a

priori that he or she has ϕ.

I believe that given any clear account of what makes a psychological property "narrow," Privileged Access logically implies the Narrow-Belief Thesis that all belief-individuating properties are narrow. I have argued at length elsewhere (McKinsey, 1991a and 1991b) that a narrow psychological property should be characterized as a psychological property whose possession does not conceptually or logically imply the existence of objects in the external world. Thus, a belief-individuating property that was *not* narrow would be one from the possession of which one *could* deduce the existence of external objects. But if such a wide property were belief-individuating, then Privileged Access implies that a person who has the property could know a priori that he has it. And then the person could simply deduce the existence of external objects from something he knows a priori, and thus the person would thereby have proven a priori that the external world exists! But since we obviously *cannot* have such a priori knowledge, it follows that belief-individuating properties must all be narrow.[7]

Thus one significant source of the internalism-externalism conflict is a perceived inconsistency between the direct reference thesis and the traditional epistemological point of view regarding the mental. The hopeful prospect offered by Modified Internalism is that, by giving up the Proposition Theory, we can consistently hold on to both the direct reference theory and Privileged Access. This hope is tempered by the sober challenge of replacing the Proposition Theory with a worthy alternative, a challenge I hope to at least partly meet below.

4. The De Re Objection.

Given adoption of the thesis of direct reference, commitment to the denial of the Narrow-Predicate Thesis seems to immediately follow, so that the main issue remaining appears to be that of adjudicating between the Narrow-Belief Thesis and the Proposition Theory. But while I think that this appearance is correct, it may not be entirely obvious that it is. In particular, some philosophers might object to the argument from direct reference on the grounds that it falsely assumes that a belief-predicate like 'believes that Bismarck was an astute diplomatist' would be a *de dicto* predicate. This is false, such a philosopher might say, because such predicates are *de re*. And this in turn might be said on the grounds that a genuine term would necessarily have large scope relative to the belief operator, so that the resulting predicate would have to be *de re*.

But the claim that a genuine term would necessarily have large scope relative to the belief operator has little to recommend it. The claim in fact seems to be self-refuting. Suppose it is said that in

(1) Russell believes that Bismarck was an astute diplomatist,

the term 'Bismarck' has large scope relative to 'believes'. Apparently, this means that (1) is semantically equivalent to

> (2) Bismarck is such that Russell believes that he was an astute diplomatist.

Defenders of a direct reference interpretation of 'Bismarck' can no doubt all agree that (1) and (2) are semantically equivalent. But now we are faced with the fact that, in order to understand the claim that 'Bismarck' has *large* scope in (1), we have to be able to assume that the anaphoric pronoun 'he' has *small* scope in (2). But of course, the anaphoric pronoun 'he' occurring in (2) is just as much a genuine term as 'Bismarck'. Thus the claim that all genuine terms must have large scope relative to the belief operator is a claim that cannot even be understood unless it is taken in such a way that it implies its own falsehood.[8]

The idea that *de dicto* belief predicates cannot be constructed from sentences that contain genuine terms therefore cannot be justified by the claim that such terms must have large scope. The only other possible justification of this idea would be to somehow argue that the *de dicto* belief operator cannot meaningfully be attached to sentences containing genuine terms, and that meaningful predicates such as 'believes that Bismarck was an astute diplomatist' are instead formed by attaching a different *de re* operator to the sentence in question. This suggestion is of course reminiscent of Quine's famous stricture that the *de dicto* (or "notional") belief operator cannot be meaningfully attached to a sentence that contains a free variable, and that meaningful belief predicates containing such a variable must instead have been constructed by use of a different *de re* (or "relational") belief operator. (See Quine, 1956.) But like Quine's stricture, the present proposal has the serious drawback of implying that the verb 'believes' is ambiguous. And this implication is clearly false. It is just preposterous to suppose, for instance, that in a *de dicto* predicate containing no genuine terms such as 'believes that the first Chancellor of a country called "Germany" was an astute diplomatist' the verb 'believes' has a different meaning than it has in the meaningful predicate 'believes that Bismarck was an astute diplomatist'.

Once we've accepted the direct reference view of names, therefore, there seems to be no plausible way of avoiding the consequence that some *de dicto* belief predicates express wide psychological properties.[9]

5. The implications of mental anaphora.

But the fact that some *de dicto* belief predicates express wide psychological properties poses a problem for the Narrow-Belief Thesis only if there is some reason to believe that the wide properties so expressed are inevitably belief-individuating. And again, the main reason for believing this is the Proposition Theory. Yet, I now wish to argue, it is precisely the cases in which the

proposition believed is wide, or singular, that make it clear that the proposition believed sometimes provides neither a complete nor an essential characterization of the belief in question.

For when a person believes a singular proposition about an external object, the person must *base* her mental reference to that object on something; the person must possess some *way* or *means* of thinking of an external object, in order to believe or think of a singular proposition involving that object. And it is arguable that this basis or means of thinking about the object must itself be part of the property that individuates the person's thought or belief. Since this feature of the belief's individuating property is something *in addition to* the singular proposition believed, the belief is not completely characterized by the proposition believed. Moreover, as we shall see, once the basis on which a person believes a given singular proposition is specified, it is in general possible to individuate the belief *independently* of the proposition believed, so that the proposition believed also fails to essentially characterize the belief.

These consequences, which refute the Proposition Theory, follow from the properties of a form of cognitive attitude ascription that I have identified elsewhere (McKinsey, 1986) and called "mental anaphora." The central type of example to consider is one in which a person mentally picks out an object on the basis of a certain property, and then wishes counterfactually that the object had not had that very property, as in

> (3) Oscar assumes that just one fish got away from him at t, and he wishes it had been the case that he caught *it* (that very fish) at t.

The point of choosing a case involving a counterfactual wish is to rule out the possibility that the anaphoric pronoun 'it' could be going proxy for a definite description recoverable from the antecedent quantifier phrase. This is ruled out in the present case, for if 'it' were short for 'the fish that got away from him at t', then the second conjunct of (3) would mean 'he wishes it had been the case that he caught at t the fish that got away from him at t', and so interpreted, (3) would ascribe a self-contradictory wish to Oscar. Yet clearly, (3) can be understood as ascribing a perfectly consistent wish, and so there is a reading of (3) on which the pronoun 'it' is not going proxy for a definite description. Moreover, it is equally clear that 'it' is also not a bound variable in (3) either, since the scope of the quantifier phrase 'just one fish' does not extend over the second conjunct of (3) in which 'it' occurs.

It seems to me that the only plausible hypothesis is that 'it' is functioning in (3) as what Evans (1977) called an "E-type pronoun." An E-type pronoun is a genuine term whose reference is fixed in Kripke's (1972) sense by the description that is recoverable from the phrase governed by the pronoun's quantifier antecedent. In (3) we use 'it' with the meaning of an E-type pronoun to ascribe the same wish that the person who has the wish would himself express by using the pronoun with the same meaning, as when Oscar says:

(4) Just one fish got away from me at t, and I wish it had been the case
     that I caught *it* (that very fish) at t.

Let us assume that Oscar is right that just one fish got away from him at t,
and let us call this fish 'Bubbles'. Then clearly, the proposition that would make
Oscar's wish come true, and the content of the wish ascribed to Oscar by (3), is
the singular proposition that Oscar catches Bubbles at t. But I think it is also
clear that Oscar's wish would not be *individuated* by this proposition. The basic
reason for this is that the wish in question is individuated in part by the
descriptive assumption on which it is based, and while this assumption
determines which proposition is wished true, by fixing an object which is one of
the proposition's constituents, the assumption is not itself a part of the
proposition wished true.

To see this, consider that the wish rather pedantically ascribed by (3) would
be more colloquially ascribed by saying simply

(5) Oscar wishes he had caught the fish that got away,

and this is intuitively a quite different wish than the one ascribed by

(6) Oscar wishes he had caught the fish that was on the end of his line.

But given that the fish that got away from Oscar *just is* the fish that was on the
end of his line, these two wishes would be made true by the very same singular
proposition, namely, the proposition that Oscar catches Bubbles (at t). For just
as (5) is correctly analyzable as (3), so (6) is analyzable as

(7) Oscar assumes that just one fish was on the end of his line at t, and
     he wishes it had been the case that he caught *it* (that very fish) at t,

and once again, given that 'it' is an E-type pronoun, the proposition that would
make the wish ascribed by (7) come true, is just the singular proposition that
Oscar catches Bubbles at t. Thus, although the wish ascribed by (3) and (5) is a
different wish from that ascribed by (6) and (7), these two different wishes have
the same propositional content. So the sameness of two wishes' propositional
contents is not a *sufficient* condition for the sameness of the wishes. In other
words, the proposition wished true does not always provide a *complete*
characterization of the wish. Hence, the Proposition Theory gives a false account
of how wishes are to be individuated.

Other implications of mental anaphora provide additional considerations
against the Proposition Theory (of wishing). Consider another possible world in
which Oscar also consistently wishes he had caught the fish that got away from
him but in which the fish that got away from Oscar is not Bubbles, but some
different fish. Intuitively, Oscar has exactly the same wish as the wish he had in
the actual world when he also wished he had caught the fish that got away from
him but the proposition wished true is a different singular proposition about a

different fish. Thus sameness of propositions wished true is not a *necessary* condition of two wishes' sameness either. In other words, the proposition wished true does not provide an *essential* characterization of the wish.

This same consequence also follows from another kind of possibility. Suppose now that the assumption on which Oscar's wish is based is false: no fish got away from him, because, say, what he had on the end of his line was not a fish at all, but an old boot, or an underwater branch. Still, it remains true that Oscar wishes he had caught the fish that got away. He has the same wish as when his wish concerned Bubbles, but in this situation his wish is not about any fish at all, and in fact there simply is *no* proposition that he is wishing true. This kind of possibility not only refutes the Propositional Theory of Individuation (for wishes), but it also refutes the Relation Theory (for wishing). For in this possible situation, the sentences (3) and (5) are true and ascribe a wish, but there is no proposition expressed by the sentence in the scope of 'wishes' (i.e., the sentence 'he caught it at t' in the case of (3)), and so contrary to the Relation Theory, the second conjunct of (3) does not ascribe the property of being mentally related to a certain proposition.[10]

6. Anaphoric Beliefs

Given the existence of mental anaphora in the case of wishing, it is clear that the same construction, with the same interpretation and similar consequences also exists in the case of all the other cognitive attitude operators. To see the effect in the case of belief, it will be useful to consider the anaphoric beliefs that may plausibly be attributed to a protagonist of the classic example of Hesperus and Phosphorus. Let us suppose that Androcles, a citizen of ancient Rome, regularly recognizes and refers to the brightest heavenly body on the western horizon in the evening. The name he uses to refer to the heavenly body so recognized and described is 'Hesperus'. Androcles also regularly recognizes and refers to the brightest heavenly body on the eastern horizon in the morning, using the name 'Phosphorus' to refer to this planet.

Now of course, Androcles does not realize that when he uses these two different names, he is referring to one and the same planet, and thus he either dissents from or withholds assent to (the Latin version of) the sentence 'Hesperus = Phosphorus'. He also, we may suppose, assents to the Latin version of 'Hesperus appears in the evening', though for most of his life, he would not have assented to 'Phosphorus appears in the evening'.

Now let us add a slight twist to the traditional story, a twist that I will use to make a point that is similar to one made by Loar (1985) and Salmon (1986) about their variations on Kripke's (1977) famous puzzle about Pierre. At some point in Androcles' life, Roman astronomers construct and confirm an exciting new theory of the heavens from which it follows that Hesperus = Phosphorus.

Androcles does not learn about this new theory at all, but one fact that follows from the theory trickles down to him (say, through an offhand remark by a better informed neighbor), namely the fact that Phosphorus also appears on the evening horizon. So now, Androcles also assents to (the Latin version of the sentence) 'Phosphorus appears in the evening', though it still does not occur to him that perhaps Hesperus *is* Phosphorus.

Androcles assents to both 'Hesperus appears in the evening' and 'Phosphorus appears in the evening'. Moreover, these two sentences express the same proposition, since, I assume, the names 'Hesperus' and 'Phosphorus' are both genuine terms that refer to the same planet, namely Venus. But the interesting question is, What *beliefs* does Androcles express when he uses these two sentences? It seems implausible to suppose that he is expressing the *same* belief when he utters the sentences. After all, when Androcles began assenting to 'Phosphorus appears in the evening', he apparently did so because he had acquired a *new* belief. It is even plausible to suppose that in so assenting, Androcles, as we say, *changed his mind*, having in the past assumed that Phosphorus does *not* appear in the evening. But since he had all along had the belief he expresses with 'Hesperus appears in the evening', the new belief expressed by 'Phosphorus appears in the evening' must be a different belief.[11]

Androcles uses two sentences that express the *same* proposition to express two *different* beliefs. Thus our example, based on those of Loar and Salmon, provides another argument against the Proposition Theory. For the example strongly suggests that Androcles' beliefs are different, even though they have the same singular proposition as content. This in turn strongly suggests that the beliefs involve *more* than just the propositions believed. But what more?

The resources of mental anaphora provide a simple answer to this question. The belief Androcles expresses by 'Hesperus appears in the evening' is the (slightly redundant) belief ascribed by

(8) Androcles assumes that just one heavenly body is brightest on the western horizon in the evening, and he believes that *it* (that very heavenly body) appears in the evening,

while the belief Androcles expresses by 'Phosphorus appears in the evening' is the (more interesting) belief ascribed by

(9) Androcles assumes that just one heavenly body is brightest on the eastern horizon in the morning, and he believes that *it* (that very heavenly body) appears in the evening.

Although the beliefs ascribed by (8) and (9) are different, the propositions believed (as opposed to those assumed) are the same (the singular proposition that Venus appears in the evening). Thus sameness of proposition believed is not *sufficient* for sameness of belief. Moreover, while Androcles would have the same belief as that ascribed by, say, (9) in other possible worlds in which (9) is

true, the propositions believed in these other worlds might not be the same. For instance, in another possible world in which (9) is true, Mars rather than Venus might be the heavenly body that is brightest on the eastern horizon in the morning. In such a world the proposition that Mars appears in the evening would be the content of the belief ascribed by (9). Or in some worlds in which (9) is true, Androcles' belief might have *no* propositional content, because, say, everyone suffers from a mass optical illusion and there is no heavenly body that is brightest on the eastern horizon in the morning. Hence, sameness of proposition believed is not *necessary* for sameness of belief, either.

As in the case of wishing, the Proposition Theory of belief is refuted by the fact that a person's way of thinking of an object can be a crucial feature of the property that individuates the person's belief, in cases where the belief's content is a singular proposition. Other philosophers, notably Loar (1985) and Salmon (1986), have also recently reemphasized the Fregean point that we need to mention a person's way of thinking of an object if we are to fully characterize the person's belief. But apparently, both Loar and Salmon also think that the resources of ordinary language, in particular the resources provided by ordinary cognitive attitude constructions, are not adequate to provide such fuller characterizations of persons' beliefs. But this is wrong. As we've seen, constructions involving mental anaphora can provide such fuller characterizations, and these constructions are of course found in ordinary language.[12]

7. The Belief-Predicate Assumption.

Our refutation of the Proposition Theory eliminates the primary reason why philosophers, internalists and externalists alike, have accepted the Belief-Predicate Assumption, which says that every property expressed by a *de dicto* belief predicate is belief-individuating. On the other hand, we have so far seen no reason to *reject* that assumption. But the reason is close at hand. We have observed that a belief cannot in general be individuated by the singular proposition believed. In fact, however, many *de dicto* belief predicates, such as 'believes that Bismarck was an astute diplomatist', express precisely the property of believing a certain singular proposition. Such predicates, therefore, do not express belief-individuating properties, and hence such predicates are counterexamples to the Belief-Predicate Assumption.

There is admittedly a tension between the claim just made about predicates like 'believes that Bismarck was an astute diplomatist' and the claims I made above about the belief predicates involved in cases of mental anaphora. The claims are consistent, however, because there are two radically different kinds of belief predicates that can be constructed by applying the operator 'believes that' to sentences containing genuine terms. This difference in turn follows from a

radical semantic difference between two types of genuine terms.

Let us just stipulate that the names 'Hesperus' and 'Phosphorus' have the kinds of descriptive meanings that are involved in the beliefs that Androcles would express by use of sentences containing these names. The meaning of 'Hesperus', let us suppose, is given by the reference fixing rule:

(10) For any token α of 'Hesperus', α is to refer to an object x if and only if x = the brightest heavenly body on the western horizon in the evening,

while the meaning of 'Phosphorus' is given by the different reference fixing rule:

(11) For any token α of 'Phosphorus', α is to refer to an object x if and only if x = the brightest heavenly body on the eastern horizon in the morning.

Given these stipulations about the meanings of 'Hesperus' and 'Phosphorus', consider what sorts of beliefs would be ascribed by the sentences

(12) Androcles believes that Hesperus appears in the evening; and
(13) Androcles believes that Phosphorus appears in the evening.

Since the names 'Hesperus' and 'Phosphorus' are by assumption both genuine terms whose reference is fixed by descriptions, the beliefs ascribed by (12) and (13) would be semantically analogous to these sentences, so that, like the sentences, the beliefs would also be singular and would also involve references that are analogously fixed by these same descriptions. Thus, if 'Hesperus' and 'Phosphorus' have the descriptive meanings in question, then (12) and (13) are respectively equivalent to the anaphoric sentences (8) and (9).[13]

But then, given what we saw above about mental anaphora, it follows that the belief contexts in (12) and (13) are classically opaque and nonrelational: since both (8) and (9) can be true in possible worlds in which no planet fits the descriptions that fix the reference of 'Hesperus' and 'Phosphorus', existential generalization on these terms in (12) and (13) is invalid; since each of (8) and (9) can be true while the other is false, even though Hesperus = Phosphorus, substitution of these coreferential genuine terms for one another in (12) and (13) is invalid.

In short, and contrary to what is usually assumed, it does not follow from the fact that a given sentence S contains a referring genuine term and hence expresses a singular proposition, that the *de dicto* predicate 'believes that S' expresses a wide property that is relational with respect to the term's referent. Belief predicates containing E-type pronouns such as the predicates occurring in (8) and (9), as well as belief predicates containing names with descriptive meanings, such as those occurring in (12) and (13), are counterexamples to the validity of such an inference.

So there was a gap in the argument against the Narrow-Predicate Thesis that

I described above, the argument which proceeded from the premise that proper names are genuine terms to the conclusion that belief predicates constructed from sentences containing proper names must express wide psychological properties. But the gap is easily repaired. If every genuine term had a descriptive linguistic meaning of the kind we stipulated 'Hesperus' and 'Phosphorus' to have, then not only would the Proposition Theory be false, but every belief predicate constructed from sentences containing genuine terms would express narrow, belief-individuating properties, and the Belief-Predicate Assumption would be true. But in fact, most genuine terms do not have such descriptive linguistic meanings. In particular, as I have argued elsewhere at some length (McKinsey, 1984), most of the ordinary proper names that we use every day have no meanings in the language at all. [14] Moreover, the other main species of genuine term, which includes indexical demonstratives and personal pronouns, also have no descriptive linguistic meanings. It follows that most *de dicto* belief predicates constructed from sentences containing ordinary genuine terms express wide, relational properties.

Consider again the predicate 'believes that Bismarck was an astute diplomatist'. If the name 'Bismarck' had a descriptive meaning fixing its reference, then it would be plausible to suppose that the name could contribute this descriptive meaning to the predicate, and thus allow the predicate to convey a believer's way of thinking about Bismarck, similar to the way in which we imagined the descriptive names 'Hesperus' and 'Phosphorus' could be used to convey two of Androcles' ways of thinking about Venus. But in fact 'Bismarck' has no descriptive meaning, for it has no linguistic meaning at all. There is thus nothing about the semantic features of this name that could allow it to convey, in a public language, anything about anyone's way of thinking about Bismarck. The only semantic property that 'Bismarck' has to contribute is its referent, and thus the only property that the *de dicto* predicate 'believes that Bismarck is an astute diplomatist' could express is the relational property of having a belief regarding Bismarck to the effect that he is an astute diplomatist. And this in fact is the intuitively correct way of understanding the predicate in question.

The same point holds for most belief predicates containing demonstratives and personal pronouns. When Alexander refers to me in the third person, saying 'Sergei believes that *he* (or: *that man*) is rich', or my wife says to me 'Sergei believes that *you* are rich', or I say 'Sergei believes that *I* am rich', all three are saying the same thing about Sergei, namely, that he has a belief regarding me to the effect that I am rich. In these cases, the pronouns cannot possibly convey anything about the way Sergei thinks of me. For while these pronouns do have linguistic meanings that have some descriptive content, the content is insufficient to characterize a believer's way of thinking about an object. This is because the meaning of such a term relies upon the context to supply part of the condition that determines the term's referent, and the context is typically independent of any way in which the *believer* might think of the referent. In fact,

in my view, the referent of such a term is typically determined in part by the *speaker's* way of thinking of the referent, and so even if this condition is somehow made clear by the context, it cannot be used to ascribe the *believer's* way of thinking.[15]

Let us say that a genuine term whose linguistic meaning is sufficient to provide a context-independent specification of the term's referent is a term that has a "context-independent" meaning. We have just observed that most genuine terms of ordinary language, including most proper names and most indexical pronouns, do not have context-independent meanings, and that as a consequence most *de dicto* belief predicates formed from sentences containing such terms express wide, or relational, belief properties. The properties in question are the same as the *de re* properties expressed by the structurally *de re* predicates with which these particular *de dicto* predicates are equivalent. Thus, as is clear from our discussion, the property expressed by the *de dicto* predicate 'believes that Bismarck was an astute diplomatist' is precisely the property expressed by the *de re* predicate 'believes of Bismarck that he was an astute diplomatist'. We might say, then, that some *structurally de dicto* predicates are *semantically de re*.

Note again that it does not follow from the fact that a given belief predicate expresses a relational property that the predicate is not (structurally) *de dicto*. Note also that even when the property expressed by a *de dicto* predicate is the same as the property expressed by a structurally *de re* predicate, the two predicates will not share all their logical properties in common. In particular, substitution of coreferential genuine terms will not in general be valid in the case of the *de dicto* predicate, though it will be valid in the case of the corresponding *de re* predicate. For instance, the premises that Androcles believes that Venus appears in the evening and that Venus = Hesperus do not logically imply that Androcles believes that Hesperus appears in the evening. For the conclusion, being by stipulation equivalent to the anaphoric sentence (8), implies that Androcles' belief is based on a certain descriptive assumption, while the premises do not imply this (I assume that 'Venus' is an ordinary name with no descriptive meaning).[16,17]

Now let us return to the Belief-Predicate Assumption. We have seen that a *de dicto* predicate like 'believes that Bismarck was an astute diplomatist' expresses a wide, relational, *de re* belief property. This in effect means that such a predicate characterizes a person's belief in terms of the singular proposition believed. But as we observed from our discussion of mental anaphora, a person's belief cannot in general be individuated in such terms. Hence, some *de dicto* belief predicates express properties that do not individuate with respect to belief, and so the Belief-Predicate Assumption is false.

But again, without the Belief-Predicate Assumption, we cannot legitimately infer from the falsity of the Narrow-Predicate Thesis that some beliefs are individuated by wide psychological properties. We may conclude, then, that the semantic facts which support a direct reference theory of ordinary names and

indexical pronouns, are perfectly consistent with the traditional view that our beliefs are individuated by narrow psychological properties. And this in turn implies that these same semantic facts are consistent with the idea that we have privileged access to the properties that individuate our beliefs.

8. The contribution problem.

My argument against the Belief-Predicate Assumption is two-pronged. It first uses the belief predicates found in constructions of mental anaphora and the equivalent predicates formed from sentences containing descriptive names to argue that beliefs are not in general individuated by the singular proposition believed. Then the argument uses the quite different belief predicates formed from sentences containing nondescriptive names and indexicals to show that some belief predicates ascribe precisely the property of believing a singular proposition and hence do not ascribe a belief-individuating property, contrary to the Belief-Predicate Assumption. I believe that the semantic facts I've cited pretty conclusively show the existence of these two radically different kinds of *de dicto* belief predicates. But the existence of these two types of belief predicates leaves us with a serious problem.

As we have seen, when a belief predicate contains an E-type pronoun in the context of mental anaphora, or contains a name with a descriptive meaning, the genuine term's referent is irrelevant to what is said by the whole belief sentence, and in fact the sentence's truth value is unaffected by whether the term has a referent or not. But when a belief predicate contains a name with no descriptive meaning, or a (non-anaphoric) indexical pronoun, then the term's referent is contributed to the proposition expressed by the whole belief sentence, and in fact the (*de re*) belief sentence will itself express a singular proposition whose very existence is dependent upon the existence of the term's referent. The problem is, How is this possible? A genuine term's sole (normal) function is to introduce a referent into what is said by the whole sentence. How is it possible that one kind of genuine term continues to contribute its referent as usual to what is said when it occurs in belief predicates, while another kind of genuine term fails to contribute its referent when it occurs in belief predicates, thus behaving abnormally? I call this "the contribution problem."[18]

I will give my solution to this problem in the form of a story, which I call "the Revised Jonesean Myth." I call the story this, since it involves a slight emendation of Sellars's (1963) famous Jonesean Myth about how the cognitive attitude predicates were introduced into ordinary language. According to Sellars, these predicates were introduced in order to provide semantic classifications of theoretical inner states. The idea is that a predicate of the form 'believes that S' is supposed to classify a person's belief as one that is semantically analogous to the sentence S. Syntactically, the operator 'believes that' was introduced on the

assumption that, for any well-formed sentence S, 'believes that S' will be a well-formed predicate.

But as far as clarity is concerned, the semantic rule for interpreting such predicates leaves a lot to be desired. The difficulty is that the sentences which occur in these predicates can differ wildly from each other in their semantics. Some of the sentences may have context-independent meanings and yet express no proposition (for the sentence might contain a descriptive genuine term that has no referent); other sentences may have no linguistic meanings at all, though they do express singular propositions (for the sentence might contain a referring proper name that has no meaning in the language); still other sentences may contain indexical pronouns and thus have context-dependent linguistic meanings, relying on context to determine which proposition is expressed. Analogies based on the semantic properties of only one of these different types of sentence are bound to fail in the case of a belief ascribed by another type of sentence, and so it is hard to see how belief predicates could all have a uniform type of interpretation. And yet we introduced the belief operator on the assumption that, for *every* well-formed sentence S, 'believes that S' is going to be a well-formed predicate that provides a way of semantically classifying a belief.

So what are these predicates supposed to mean? The fact is that when we introduced belief predicates into our language we didn't know what we were doing. For the predicates are in effect *metalinguistic* and semantic in nature, and yet we did not then, nor do we now, know much of anything about the semantics of the sentences on the basis of which the "analogous" belief properties are supposed to be constructed. Yet we went ahead and introduced the predicates anyway, and let the semantic properties of the sentences of our language (whatever they may be) dictate what the resulting belief predicates might say about our beliefs. We introduced the predicates, and let the semantic chips fall where they may.

And this is what happened. When we introduced the belief predicates, we *wanted* to be able to use them to say as much as possible about the semantic properties of our beliefs. In short, we *hoped* that the predicates would end up expressing belief-individuating properties. And so we gave this interpretation priority over all others. If the sentence S in the predicate 'believes that S' has sufficient semantic characteristics to fully characterize a corresponding belief, then the predicate is interpreted as expressing an individuating property. This means in particular that when S contains a descriptive genuine term, the referent of the term (if any) drops out as irrelevant to the classification of the belief; for the referent will be irrelevant to any property that might individuate the belief. Since this is the kind of interpretation that a belief predicate is always given if it is at all possible to do so, I will call it the "default" interpretation.

But the default interpretation does not always work, as our linguistic intuitions dimly tell us. In particular, when the sentence S in 'believes that S' contains a proper name with no descriptive meaning, or contains a (non-

anaphoric) indexical with a context dependent meaning, S is too semantically poverty stricken to allow application of the default interpretation. In such a case, the genuine term's referent becomes relevant to the interpretation, for without it there would *be* no interpretation. If in such cases, we did not use the referent (and hence the singular proposition expressed) to semantically classify the belief, we could not use S to classify the belief at all. So when the default fails, we use the term's referent to classify the belief, and the predicate becomes semantically *de re*.

In short, the quite different kinds of semantic properties possessed by the sentences of our language force us to construct quite different kinds of belief predicates out of these sentences. As a result, there lies within the class of structurally *de dicto* predicates a sharp distinction between those that are semantically *de re* and those that are semantically *de dicto*, that is, those predicates that express narrow, individuating belief properties that are not *de re*. If this is right, then the traditional semantic distinction between *de re* and *de dicto* exists within the class of structurally *de dicto* belief predicates. Moreover, and contrary to Quine (1956), the distinction can be described without having to suppose that the verb 'believes' ever has more than one meaning. For the radical difference between semantically *de re* and *de dicto* belief predicates is not caused by using 'believes' in more than one sense when forming the predicates; rather, the difference is caused by the different ways of classifying beliefs that inevitably result from applying a single belief operator to sentences that radically differ from each other in their semantic properties.

9. Belief predicates containing kind terms.

One of the most interesting types of *de dicto* belief predicate that express wide psychological properties was discovered by Burge (1982) as a previously unnoticed application of Putnam's (1975) famous Twin Earth example. Oscar, a resident of Earth, believes that water is wet. On Twin Earth, there is no water; rather, there is a qualitatively identical liquid with a different chemical composition, a liquid that we may call 'twater'. Toscar, a resident of Twin Earth who is Oscar's identical twin, would apparently not believe that water is wet. For Toscar has no beliefs about water at all; rather, he believes that twater is wet, that twater flows from faucets, etc. Yet Oscar and Toscar, being twins, would apparently share all their narrow states. Thus, Burge concludes, Oscar's belief that water is wet must be a wide state; it must, that is, "presuppose" or "depend upon" the relations that Oscar bears to other speakers or objects in his environment.

I agree that the property expressed by the predicate 'believes that water is wet' is a wide property. But I disagree with Burge's further conclusion that the belief ascribed by use of this predicate would also be wide, that is, would be

individuated by the wide property expressed by the predicate. For in reaching this conclusion, Burge has implicitly made the Belief-Predicate Assumption, and in my view it is precisely the belief predicates which express wide psychological properties, like 'believes that water is wet', that provide the best reason for *rejecting* the Belief-Predicate Assumption.

Let us consider in detail why belief predicates containing natural kind terms express wide properties. The explanation is interestingly different than in the case of belief predicates containing names and indexicals. The latter kinds of predicates are typically wide because most ordinary names and indexicals have no context-independent linguistic meanings. As a result the terms must contribute their referents to the predicates, and thus the predicates must express *de re* belief properties. But this explanation won't work in the case of belief predicates that contain natural kind terms. For natural kind terms are surely not indexicals, and unlike proper names, they surely do have linguistic meanings in the public language. Thus these meanings, being context-independent, should be usable to characterize a believer's way of thinking about water. Moreover, it is counterintuitive to suppose that a predicate like 'believes that water is wet' would be *de re* with respect to the semantic contribution of the term 'water'. For surely, it is possible to believe that water is wet even though it should turn out that there is no such thing as water, and no natural kind to which such stuff belongs.

What then can account for the wideness of the property expressed by a predicate like 'believes that water is wet'? I have proposed elsewhere that a predicate like 'is water' is used to ascribe the purely relational property of belonging to a certain natural kind, and that the linguistic meaning of 'water' is given by a semantic rule that determines, or fixes, the property in question (McKinsey, 1987). But the Twin Earth example shows that the relevant semantic rule cannot be specified in purely qualitative or conceptual terms, since then there would be no way to distinguish the meaning of 'water' as we use it from the meaning of 'water' as used by our counterparts on Twin Earth. Thus, I proposed, the linguistic meaning of 'water' must be *objectual*. Like a singular proposition, the semantic rule for 'water' must essentially involve a contingent object that will serve to distinguish our rule from the Twin Earth rule. I proposed the following as a plausible candidate for the objectual rule that governs 'water':

> (14) For any Φ, if Φ is a token of 'is water', then for any property F, Φ is to predicate F if and only if: there is just one natural kind K such that in the actual world, liquid that is Q and that *we* have experienced belongs to K, and·F = the property of belonging to K.

Here, I let 'Q' abbreviate a set of qualities that are commonly associated with 'water', such as those of being odorless, wet, thirst-quenching, tasteless, etc. As used by a speaker on Earth, the pronoun 'we' would be used in (14) to

demonstratively pick out a group consisting of the speaker and other inhabitants of Earth, so that the rule expressed would be an objectual rule that essentially involves that group. Such a speaker's twin on Twin Earth, by contrast, would use 'we' to pick out a group of Twin Earth's inhabitants, so that the rule expressed and followed on Twin Earth, though qualitatively identical with the Earthian rule, would be a different rule involving a different group.

Now consider what sort of property would be ascribed to Oscar by the predicate 'believes that water is wet', given this theory of the meaning of 'water'. Since the meaning is given by a descriptive rule that fixes the property predicated by 'is water', the belief predicate can be used to convey part of the way in which Oscar thinks of water. Using mental anaphora, we can capture the belief property ascribed to Oscar as follows:

> (15) Oscar assumes that there is just one natural kind to which liquid
>   that is Q and that *we* have experienced belongs, and Oscar believes
>   that all stuff belonging to *that kind* is wet.

Notice that on this account, both existential generalization and substitution of coextensive terms would be invalid with respect to the occurrence of 'water' in the predicate 'believes that water is wet'. Thus the predicate is not semantically *de re* with respect to 'water'. Nevertheless, the predicate would not fully characterize Oscar's belief, because the predicate is *de re* with respect to the group implicitly referred to in the linguistic meaning of 'water'. Thus the predicate 'believes that water is wet' contains an implicit *de re* element, so that the property it expresses is wide, and although it comes close, the property does not completely succeed in individuating the subject's belief.

To do that, we would have to characterize the way in which Oscar thinks of the group in question. Suppose he picks out the group as "the group consisting of myself and all who bear R to me," for some relation R. Then the full characterization of Oscar's belief, using mental anaphora, would go as follows:

> (16) Oscar assumes that there is just one group consisting of himself
>   and all who bear R to him, Oscar assumes that there is just one
>   natural kind to which liquid that is Q and experienced by *that group*
>   belongs, and Oscar believes that all stuff belonging to *that kind* is
>   wet.

I believe that there is a clear sense in which the sort of complex mental property ascribed by an instance of (16) would be narrow: the fact that Oscar has such a property would not logically imply the existence of any objects external to Oscar; it would be a property that Oscar could in principle possess in a solipsistic world.[19] Thus, as we saw in the case of belief predicates containing ordinary names and indexicals, belief predicates containing natural kind terms express wide properties that do not fully characterize the beliefs that make the predicates true of objects. To individuate the beliefs in question, one must

instead use predicates that express narrow properties.

The general strategy I am proposing for dealing with *de dicto* belief predicates that express wide psychological properties should by now be clear. The properties expressed by such predicates are wide because the predicates contain some *de re* element that makes the belief property expressed relational with respect to some contingent object or substance. But then, since the property characterizes a belief in terms of some object, the property cannot be used to individuate a belief. To do that, the property would have to characterize the person's way of thinking of the object in question. Ascription of belief properties that succeed in individuating a person's belief by expressing the person's way of thinking about an object can be accomplished by use of other *de dicto* belief predicates that utilize mental anaphora to ascribe narrow psychological properties.[20]

In this paper, I have tried to show that direct reference theories of proper names, indexical pronouns, and natural kind terms are consistent with the traditional views that we each have privileged access to the properties that individuate our beliefs and other cognitive attitudes, and that these individuating properties are consequently all narrow, not logically implying the existence of external objects. Crucial to my strategy was my argument against the Belief-Predicate Assumption to the effect that the wide properties frequently expressed by *de dicto* belief predicates cannot completely or essentially characterize a person's beliefs, and my proposal that in such cases, we can use ordinary belief constructions involving mental anaphora plus predicates with narrow meanings to individuate the person's belief.[21]

Notes

1. Brian Loar (1985) was, I believe, the first to clearly expose and cogently argue against this assumption.
2. We shall see below that sometimes, a given belief property may be possessed by more than one of a person's belief states. This is why I stipulate here that the belief state is *a* state of x's being F, rather than *the* state of x's being F.
3. In this definition, I am assuming that properties that individuate beliefs do so *across* possible worlds, so that a belief with a given individuating property in one possible world will be the same belief as a belief in another possible world just in case the latter belief also has the property in question. I believe that this assumption is intuitively correct. Certainly, the assumption is commonly taken for granted in discussions of belief-individuation. See, for instance, Burge (1979) and Loar (1985).
4. For the sake of accuracy, I should note that Frege and Russell held the Narrow-Belief and Narrow-Predicate Theses in slightly different senses of 'narrow'. For Frege, the properties expressed by *de dicto* belief-predicates are purely conceptual or qualitative. We might say that these properties are "narrow" in Putnam's (1975) original sense, on which no narrow psychological property is relational with respect to any contingent object. Russell, by contrast, allowed certain contingently relational properties expressed by belief-predicates to also individuate beliefs, provided that the *relata* in the properties were objects of

immediate mental acquaintance such as selves, mental acts or states, and sense-data. From here on, I will intend the expression "narrow" to apply to these Russellian properties as well as the Fregean ones, and so I will be using the expression in a slightly wider sense than Putnam's. For motivation and detailed analyses of these two senses of 'narrow', see McKinsey (1991a).

5. Russell (1912) was clearly motivated by epistemological considerations when he used his Principle of Acquaintance to argue against the direct reference view of ordinary names. Unfortunately, Russell did a poor job of motivating his principle, so that his argument has not received the credit it deserves. The argument I described above against direct reference theories that is based on the Narrow-Belief Thesis and the Proposition Theory, with the Narrow-Belief Thesis in turn supported by the Principle of Privileged Access given below, provides, I believe, an accurate reconstruction of the considerations underlying Russell's Principle of Acquaintance, and serves to show how powerful his argument really was.

6. For fuller discussion of this sense of a priori knowledge, see McKinsey (1987).

7. I gave this argument in (1991b). I first raised Privileged Access as a problem for direct reference and causal theories of names in my doctoral dissertation (Indiana University, 1976) and also in my (1978a). Recently the conflict between Privileged Access and externalistic views of the mental has been discussed by LePore and Loewer (1986), Davidson (1987) and Burge (1988). All of these philosophers claim that there is no inconsistency between Privileged Access and the denial of the Narrow-Belief Thesis. I believe that my argument in this paragraph and in (1991b) makes clear that the inconsistency is real. Those who harbor any hope of avoiding the inconsistency, I believe, do so largely because of unclarity about the concepts of narrow and wide psychological states.

8. I think I was first convinced of this by reading Salmon (1986) and especially Kaplan (1986).

9. As we shall see in Section 7 below, there is an important consideration that can block this inference as I've stated it. However, an additional premise will justify this result in the end.

10. An objection to my argument concerning mental anaphora was recently raised by Steven Boër (1989). He points out that in a construction like (3), the anaphoric pronoun might be short for a *rigid* definite description recoverable from the quantifier antecedent. Although Boër does not say so explicitly, his words suggest that his proposal is that the pronoun 'it' in (3) should be treated as short for the "world-indexed" rigid description 'the fish that got away from Oscar at t in α', where 'α' is understood to be a proper name of the actual world. (On the idea of world-indexed properties and descriptions, see Plantinga (1974) and (1978).) On this proposal, the second conjunct of (3) would read: 'Oscar wishes it had been the case that he caught at t the fish that got away from him at t in α'. Note that we may rewrite this proposal as follows:

> (i) Oscar wishes that the following had been the case: there is an object x such that x, and only x, is a fish that got away from him at t in α, and he caught x at t.

Now it is true that (i) has the virtue of ascribing to Oscar a wish whose propositional content is perfectly consistent. The proposition wished true will be true at a possible world w just in case Oscar catches at t in w the fish that got away from Oscar at t in α, and this is certainly a possible circumstance. However, as Boër seems to anticipate, I would object that (i) distorts the content of Oscar's wish. On Boër's proposal, the content of Oscar's wish is a proposition whose truth logically implies that just one fish got away from Oscar at t in α. But this, in my view, is phenomenologically wrong. Oscar is not wishing for something to occur that logically requires a fish to have actually gotten away from him. Rather, he is wishing that something had occurred which he assumes did not in fact occur at all. On my view, unlike Boër's, the assumption that just one fish

actually got away from Oscar is not a part of the fulfillment condition of Oscar's wish. Rather, the assumption provides the basis on which Oscar is able to *think* of the fulfillment condition of his wish (i.e., the proposition that he catches a certain fish at t).

There is another serious problem with Boër's proposal, a problem deriving from the fact that the proposal in effect replaces direct reference in Oscar's wish to the fish that got away by direct reference to the actual world α. Thus the proposal requires an independent explanation of how Oscar pulls off this direct reference to α. Now the only mechanism for mentally referring directly to the actual world that suggests itself is that this mental reference is fixed by the description 'the actual world'. Thus, if Oscar's wish (as ascribed by (i)) involves direct reference to the actual world, then this would seem to be possible only because that wish in turn is based upon a further descriptive assumption, which we would explicitly ascribe by saying

> (ii) Oscar assumes that there is just one actual world, and Oscar wishes that the following had been the case: there is an object x such that x, and only x, is a fish that got away from him at t in *it* (that very world) and he caught x at t.

Now of course we cannot, on pain of circularity, suppose that the pronoun 'it' in (ii) is short for the world-indexed description 'the world that is actual in α', and hence we must suppose that 'it' is an E-type pronoun whose reference is fixed by the description 'the actual world'. Thus it certainly appears that in order to make sense of Boër's proposal, the mechanism of anaphoric reference described in my initial proposal must be assumed to exist anyway, and so the detour through Boër's proposal is both unnecessary and unmotivated.

It might be claimed that my proposal of (ii) as a fuller characterization of the wish ascribed by (i) is wrong, because the wish ascribed by (i) is supposed to be (somehow) irreducibly *de re* with respect to the actual world α. But this claim has the absurd consequence that in any possible world *other than* α, Oscar could not wish that he had caught the fish that got away. For in any possible world w other than α, Oscar can have no epistemic contact with α; in w, Oscar can have epistemic contact only with the possible world that is actual *relative to w*, namely, w itself. Thus in w, Oscar will only be able to mentally refer to w, and so in w he can have no wish that is about α at all. Thus if (i) ascribed a wish that is irreducibly about α, then (i) could not possibly serve as a correct account of Oscar's wish; for there are obviously many possible worlds besides α in which Oscar can wish that he'd caught at t the fish that got away from him at t. In order for Boër's proposal to do justice to this fact, it is necessary to suppose that (ii) more fully characterizes what, on that proposal, is Oscar's wish. (Note that (ii) can be true in possible worlds distinct from α.) So again, Boër's own proposal is committed to the existence of the very type of anaphoric reference that is described in my initial proposal. (I am pretty sure that my argument in this paragraph was inspired by a similar point that G.W. Fitch (1981, p.30) made against the view that proper names are short for world-indexed descriptions.)

11. Compare Salmon's discussion of his puzzle about Elmer's Befuddlement, in his (1986), pp. 95-98.

12. The consequences of mental anaphora lead to a different overall point of view toward the Proposition Theory than is usually endorsed, even by those who correctly emphasize ways of thinking as crucial to the individuation of singular beliefs. For instance, Perry (1979), Fitch (1985) and Salmon (1986) all seem to endorse views on which *both* the singular proposition believed *and* the person's way of believing it are crucial to the belief's individuation. Their views, then, correctly reject sameness of proposition believed as *sufficient* for sameness of belief, but incorrectly accept sameness of proposition believed as *necessary* for sameness of belief. Mental anaphora shows that sameness of proposition believed is neither necessary nor sufficient for sameness of belief, for it shows

that once a person's way of thinking of a singular proposition is provided, the person's belief can be individuated *independently* of that proposition. Similarly, in continuing to endorse the Relation Theory of cognitive attitude verbs, Salmon (1986) is also retaining another traditional feature of the Proposition Theory which mental anaphora shows is false.

13. See McKinsey (1986), p.172-173.

14. In my (1984), I argued at length against causal theories of names and defended a form of description theory according to which most ordinary names are genuine terms whose referents are determined by private, descriptive, semantic rules. Since the rules are private, they cannot give meanings to names in actual languages.

15. For my views on indexical and demonstrative reference see McKinsey (1979), (1983a), and (1984).

16. For helping me see a need for a correction in the above paragraph, I am indebted to David Shier (1993), Chapter 2. So far, we have only observed that there are *possible* cases of substitution failure of coreferential genuine terms. There are in my view *actual* such cases only if there really are some genuine terms that have descriptive meanings in the language. I think that there probably are such terms, but they are also probably rare. Consider 'Jack the Ripper', for instance. (The example of 'Jack the Ripper' as a descriptive name was proposed by Kripke (1972), p. 291.). Henry has a quiet and self effacing neighbor named Percy Smythe, and Henry believes that Percy Smythe would never harm anyone. Unbeknownst to Henry of course Percy Smythe is the notorious Jack the Ripper. Does Henry believe that Jack the Ripper would never harm anyone? Clearly not, and the failure of substitution here seems due to the fact that, unlike most names, 'Jack the Ripper' has a descriptive meaning. Many problems remain. Do the merely possible existence of descriptive names like 'Hesperus' and 'Phosphorus' plus the actual existence of a few descriptive names like 'Jack the Ripper' suffice to adequately account for all apparent substitution failures involving genuine terms? I believe so, but I will have to leave discussion of this difficult matter to another occasion.

17. My account of *de dicto* belief predicates containing proper names provides, I believe, a satisfactory solution to Kripke's (1977) famous puzzle about belief. At one point in his life, Androcles might have appeared to have contradictory beliefs, for he both believed that Venus appears in the evening (since he believed that Hesperus so appears, and Hesperus exists and is Venus) and also believed that Venus does *not* appear in the evening (since he believed that Phosphorus does not so appear, and Phosphorus exists and is Venus). But contrary to appearances, Androcles never had inconsistent beliefs. Since the name 'Venus' is nondescriptive, the structurally *de dicto* predicates 'believes that Venus appears in the evening' and 'believes that Venus does not appear in the evening' are both semantically *de re* and only partly characterize the beliefs that make the predicates true of Androcles. When these beliefs are fully characterized (either by use of mental anaphora or by use of descriptive names), with Androcles' two distinct ways of thinking about Venus made explicit, the rationality of Androcles' belief states becomes apparent.

    Essentially the same solution to Kripke's puzzle was proposed by Salmon (1986), pp. 129-132. Notice that my version of Kripke's puzzle is constructed without use of any of the principles about language (of disquotation and translation) that Kripke used in stating his puzzle. That this can be done, making clear that the puzzle is purely a puzzle about *belief*, was persuasively pointed out by Donnellan (1989).

18. I first became clearly aware of this problem as the result of a question that Ray Elugardo asked me about mental anaphora in correspondence.

19. In McKinsey (1991a) I apply mental anaphora in a similar way to show how it is possible for a person to grasp the public, objectual meaning of a natural kind term like 'water' in a private, wholly internal, narrow mental state.

20. This general strategy of course assumes that when a person satisfies a belief predicate containing words with wide content like 'Bismarck' or 'water', we can in typical cases more fully characterize the person's belief by utilizing *other* belief predicates that contain only words with narrow content. But defenders of externalism often object to this kind of strategy by claiming that there will just not be *enough* words with narrow content to do the job. Perhaps the most popular way of justifying this claim is by appeal to Burge's (1979) famous 'arthritis'-argument. Both internalists such as Fodor (1987) and externalists such as Stalnaker (1989) seem to believe that Burge's argument shows that Putnam's result regarding natural kind terms applies to all general terms whatever, and hence shows that all such terms have wide meaning. However, I have shown elsewhere (in McKinsey, 1993) that even if Burge's description of his basic thought experiment is completely correct, his argument does not really come close to supporting this conclusion.

    Another popular way of justifying the claim that no general terms have narrow content is by appeal to the further claim that the correct account of every general term's meaning must be given by a "causal theory," according to which the word's meaning is determined by speakers' causal relations to external objects. I think that many philosophers believe that Kripke's and Putnam's work on proper names and natural kind terms "proves" that the correct semantics of these kinds of words must be given in a causal theory of some sort. And many other philosophers have apparently gone on to infer that the reference or meaning of *all* words are determined by causal relations to external objects, and even that the very concepts of reference and meaning themselves must be explicated in causal terms. (See, for instance, Devitt (1981) and Stalnaker (1984).) But I am unmoved by such grandiose claims. For first, I believe that I have shown in my own work that Kripke and Putnam have not in fact proved that either the reference or meaning of either proper names or natural kind terms are determined by causal relations to external objects. (See McKinsey 1978a, 1978b, 1981, 1983b, 1984, 1987, and 1991a). And second, even if Kripke and Putnam *had* proved that names and natural kind terms satisfy causal theories, it would be wrong to infer that the same result had been "proved" for every other word. Thus it would also be wrong to argue against my view that many general terms have narrow content on the grounds that Kripke and Putnam had already "proved" otherwise. (McGinn (1989) has recently argued persuasively that many types of general terms do not satisfy a causal theory of meaning, since Putnam's twin earth case does not apply to them.)

21. This paper was written while I was on sabbatical leave from Wayne State University and serving as a Fulbright Scholar in Russia during the academic year 1991-92. I am most grateful to Wayne State University, the Fulbright Program, and Moscow M.V. Lomonosov State University for their support of this research. An earlier draft of the paper was presented to the Departments of Philosophy of the Universities of Uppsala, Stockholm, and Helsinki in February, 1992, and to the Institute of Philosophy of the Russian Academy of Sciences, Moscow (June, 1992). I am grateful to participants on those occasions for useful discussions. I am also grateful to the colleagues and students who attended two of my seminars (in Winter, 1988 and Fall, 1990), had these ideas inflicted upon them when the ideas were in a particularly inchoate stage, and helped me think things through. I especially thank Barbara Humphries, Lawrence Lombard, Mitch Rognstad, Bruce Russell, David Shier, and Paul Wagoner.

References

Boër, Steven E.: 1989, "Neo-Fregean Thoughts," *Philosophical Perspectives, 3, Philosophy of Mind and Action Theory, 1989*, 187-224.

Burge, Tyler: 1979, "Individualism and the Mental," *Midwest Studies in Philosophy*

4, 73-121.

Burge, Tyler: 1982, "Other Bodies," in A. Woodfield (ed.), *Thought and Object: Essays in Intentionality* (Oxford: The Clarendon Press), 97-120.

Burge, Tyler: 1988, "Individualism and Self-Knowledge," *Journal of Philosophy* 85, 649-663.

Davidson, Donald: 1987, "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association* 60, 441-458.

Devitt, Michael: 1981, *Designation* (New York: Columbia University Press).

Donnellan, Keith: 1970, "Proper Names and Definite Descriptions," *Synthese* 21, 335-358.

Donnellan, Keith: 1974, "Speaking of Nothing," *Philosophical Review* 83, 3-31.

Donnellan, Keith: 1989, "Belief and the Identity of Reference," *Midwest Studies in Philosophy* 14, 275-288.

Evans, Gareth: 1977, "Pronouns, Quantifiers, and Relative Clauses (I)," *Canadian Journal of Philosophy* 7, 467-536.

Evans, Gareth: 1982, *Varieties of Reference* (Oxford: The Oxford University Press).

Fitch, G.W.: 1981, "Names and the 'De Re-De Dicto' Distinction," *Philosophical Studies* 39, 25-35.

Fitch, G.W.: 1985, "On the Logic of Belief," *Nous* 19, 205-228.

Fodor, Jerry: 1987, *Psychosemantics* (Cambridge, Mass.: The MIT Press).

Frege, Gottlob: 1892, "On Sense and Reference," in *Translations from the Writings of Gottlob Frege*, P. Geach and M. Black (eds.), (Oxford: The Oxford University Press, 1966).

Kaplan, David: 1968, "Quantifying In," *Synthese* 19, 178-214.

Kaplan, David: 1979, "On the Logic of Demonstratives," in P. French, T. Uehling, and H. Wettstein (eds.), *Contemporary Perspectives in the Philosophy of Language* (Minneapolis: University of Minnesota Press), 401-412.

Kaplan, David: 1986, "Opacity," in L. Hahn and P. Schilpp (eds.), *The Philosophy of W.V. Quine* (La Salle, Illinois: Open Court), 229-289.

Kripke, Saul: 1972, "Naming and Necessity," in D. Davidson and G. Harman (eds.), *Semantics of Natural Language* (Dordrecht: D. Reidel), 253-355.

Kripke, Saul: 1977, "A Puzzle About Belief," in A. Margalit (ed.), *Meaning and Use* (Dordrecht: D. Reidel), 239-283.

LePore, Ernest, and Loewer, Barry: 1986, "Solipsistic Semantics," *Midwest Studies in Philosophy* 10, 595-614.

Loar, Brian: 1985, "Social Content and Psychological Content," presented to the twenty-fifth annual Oberlin Colloquium in Philosophy; in R. Grimm and D. Merrill (eds.), *Contents of Thoughts* (Tucson: The University of Arizona Press), 99-110.

McDowell, John: 1977, "On the Sense and Reference of a Proper Name," *Mind* 86, 159-185.

McGinn, Colin: 1989, *Mental Content* (Oxford: Basil Blackwell).

McKinsey, Michael: 1978a, "Names and Intentionality," *Philosophical Review* 87, 171-200.

McKinsey, Michael: 1978b, "Kripke's Objections to Description Theories of Names," *Canadian Journal of Philosophy* 8, 485-497.

McKinsey, Michael: 1979, "The Ambiguity of Definite Descriptions," *Theoria* 45, 78-89.

McKinsey, Michael: 1981, "Causes and Intentions: A Reply," *Philosophical Review* 90, 408-423.

McKinsey, Michael: 1983a, "Psychologism in Semantics," *Canadian Journal of Philosophy* 13, 1-25.

McKinsey, Michael: 1983b, "Review of *Designation*, by Michael Devitt," *Canadian Philosophical Reviews* 3, 112-116.

McKinsey, Michael: 1984, "Causality and the Paradox of Names," *Midwest Studies in Philosophy* 9, 491-415.

McKinsey, Michael: 1986, "Mental Anaphora," *Synthese* 66, 159-175.

McKinsey, Michael: 1987, "Apriorism in the Philosophy of Language," *Philosophical Studies* 52, 1-32.

McKinsey, Michael: 1991a, "The Internal Basis of Meaning," *Pacific Philosophical Quarterly* 72, 143-169.

McKinsey, Michael: 1991b, "Anti-Individualism and Privileged Access," *Analysis* 51, 9-16.

McKinsey, Michael: 1993, "Curing Folk Psychology of `Arthritis'," *Philosophical Studies*, 70, 49-63.

Perry, John: 1979, "The Problem of the Essential Indexical," *Nous* 13, 3-21.

Plantinga, Alvin: 1974, *The Nature of Necessity* (Oxford: The Clarendon Press).

Plantinga, Alvin: 1978, "The Boethian Compromise," *American Philosophical Quarterly* 15, 129-138.

Putnam, Hilary: 1970, "Is Semantics Possible?" in S. Schwartz (ed.), *Naming, Necessity, and Natural Kinds* (Ithaca: Cornell University Press, 1977).

Putnam, Hilary: 1975, "The Meaning of 'Meaning'," in K. Gundarson (ed.), *Minnesota Studies in the Philosophy of Science* 7 (Minneapolis: The University of Minnesota Press), 131-193.

Quine, W. V: 1956, "Quantifiers and Propositional Attitudes," *Journal of Philosophy*, 53. Reprinted in W. V. Quine, *The Ways of Paradox* (New York: Random House, 1966).

Russell, Bertrand: 1912, *The Problems of Philosophy* (New York: Oxford University Press, 1959).

Salmon, Nathan: 1986, *Frege" Puzzle* (Ridgeview Publishing Co.).

Sellars, Wilfrid: 1963, "Empiricism and the Philosophy of Mind," in W. Sellars, *Science, Perception and Reality* (Ridgeview Publishing Co.).

Shier, David: 1993, "Russellian Non-Parallelism," Ph.D. Dissertation, Wayne State University.

Stalnaker, Robert C.: 1984, *Inquiry* (Cambridge, Mass.: The MIT Press).

Stalnaker, Robert C.: 1989, "On What's in the Head," *Philosophical Perspectives 3, Philosophy of Mind and Action Theory, 1989*, 287-316.

White, Stephen: 1982, "Partial Character and the Language of Thought," *Pacific Philosophical Quarterly* 63, 347-365.