Ontological Promiscuity

Jerry R. Hobbs
Artificial Intelligence Center
SRI International
and
Center for the Study of Language and Information
Stanford University

Abstract

To facilitate work in discourse interpretation, the logical form of English sentences should be both close to English and syntactically simple. In this paper I propose a logical notation which is first-order and nonintensional, and for which semantic translation can be naively compositional. The key move is to expand what kinds of entities one allows in one's ontology, rather than complicating the logical notation, the logical form of sentences, or the semantic translation process. Three classical problems – opaque adverbials, the distinction between *de re* and *de dicto* belief reports, and the problem of identity in intensional contexts – are examined for the difficulties they pose for this logical notation, and it is shown that the difficulties can be overcome. The paper closes with a statement about the view of semantics that is presupposed by this approach.

1 Motivation

The *real* problem in natural language processing is the interpretation of discourse. Therefore, the other aspects of the total process should be in the service of discourse interpretation. This includes the semantic translation of sentences into a logical form, and indeed the logical notation itself. Discourse interpretation processes, as I see them, are inferential processes that manipulate or perform deductions on logical expressions encoding the information in the text and on other logical expressions encoding the speaker's and hearer's background knowledge. These considerations lead to two principal criteria for a logical notation.

Criterion I: The notation should be as close to English as possible. This makes it easier to specify the rules for translation between English and the formal language, and also makes it easier to encode in logical notation facts we normally think of in English. The ideal choice by this criterion is English itself, but it fails monumentally on the second criterion.

Criterion II: The notation should be syntactically simple. Since discourse processes are to be defined primarily in terms of manipulations performed on expressions in the logical notation, the simpler that notation, the easier it will be to define the discourse operations.

The development of such a logical notation is usually taken to be a very hard problem. I believe this is because researchers have imposed upon themselves several additional constraints – to adhere to stringent ontological scruples, to explain a number of mysterious syntactic facts as a by-product of the notation, and to encode efficient deduction techniques in the notation. Most representational difficulties go away if one rejects these constraints, and there are good reasons for rejecting each of the constraints.

Ontological scruples: Researchers in philosophy and linguistics have typically restricted themselves to very few (although a strange assortment of) kinds of entities – physical objects, numbers, sets, times, possible worlds, propositions, events, and situations – and all of these but the first have been controversial. Quine has been the greatest exponent of ontological chastity. His argument is that in any scientific theory, "we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which the disordered fragments of our experience can be fitted and arranged." (Quine, 1953, p. 16.) But he goes on to say that "simplicity ... is not a clear and unambiguous idea; and it is quite capable of presenting a double or multiple standard." (*Ibid.*, p. 17.) Minimizing kinds of entities is not the only way to achieve simplicity in a theory. The aim in this enterprise is to achieve simplicity by minimizing the complexity of the *rules* in the system. It turns out this can be achieved by multiplying kinds of entities, by allowing as an entity everything that can be referred to by a noun phrase.

Syntactic explanation: The argument here is easy. It would be pleasant if an explanation of, say, the syntactic behavior of count nouns and mass nouns fell out of our underlying ontological structure at no extra cost, but if the extra cost is great complication in statements of discourse operations, it would be quite unpleasant. In constructing a theory of discourse interpretation, it doesn't make sense for us to tie our hands by requiring syntactic explanations as well. The problem of discourse is at least an order of magnitude harder than the problem of syntax, and syntax shouldn't be in the driver's seat.

Efficient deduction: There is a long tradition in artificial intelligence of building control information into the notation, and indeed much work in knowledge representation is driven by this consideration. Semantic networks and other notational systems built around hierarchies (Quillian, 1968; Simmons, 1973; Hendrix, 1975) implicitly assign a low cost to certain types of syllogistic reasoning. The KL-ONE representation language (Schmolze and Brachman, 1982) has a variety of notational devices, each with an associated efficient deduction procedure. Hayes (1979) has argued that frame representations (Minsky, 1975; Bobrow and Winograd, 1977) should be viewed as sets of predicate calculus axioms together with a control component for drawing certain kinds of inferences quickly. In quite a different vein, Moore (1980) uses a possible worlds notation to model knowledge and action in part to avoid inefficiencies in theorem-proving.

By contrast, I would argue against building efficiencies into the notation. From a psychological point of view, this allows us to abstract away from the details of implementation on a particular computational device, increasing the generality of the theory. From a technological point of view, it reflects a belief that we must first determine empirically the most common classes of inferences required for discourse processing and only then seek algorithms for optimizing them.

In this paper I propose a flat logical notation with an ontologically promiscuous semantics. One's first naive guess as to how to represent a simple sentence like

A boy builds a boat.

is as follows:

$$(\exists x, y) build(x, y) \wedge boy(x) \wedge boat(y)$$

This simple approach seems to break down when we encounter the more difficult phenomena of natural language, like tense, intensional contexts, and adverbials, as in the sentence

A boy wanted to build a boat quickly.

These phenomena have led students of language to introduce significant complications in their logical notations for representing sentences. My approach will be to maintain the syntactic simplicity of the logical notation and expand the theory of the world implicit in the semantics to accommodate this simplicity. The representation of the above sentence, as is justified below, is

$$(\exists e_1, e_2, e_3, x, y) Past(e_1) \wedge want'(e_1, x, e_2) \wedge quick'(e_2, e_3) \\ \wedge build'(e_3, x, y) \wedge boy(x) \wedge boat(y)$$

That is, e_1 occurred in the past, where e_1 is x's wanting e_2, which is the quickness of e_3, which is x's building of y, where x is a boy and y is a boat.

In brief, the logical form of natural language sentences will be a conjunction of atomic predications in which all variables are existentially quantified with the widest possible scope. Predicates will be identical or nearly identical to natural language morphemes. There will be no functions, functionals, nested quantifiers, disjunctions, negations, or modal or intensional operators.

2 The Logical Notation

Davidson (1967) proposed a treatment of action sentences in which events are treated as individuals. This facilitated the representation of sentences with time and place adverbials. Thus we can view the sentences

John ran on Monday.
John ran in San Francisco.

as asserting the existence of a running event by John and asserting a relation between the event and Monday or San Francisco. We can similarly view the sentence

John ran slowly.

as expressing an attribute about a running event. Treating events as individuals is also useful because they can be arguments of statements about causes:

Because he wanted to get there first, John ran.
Because John ran, he arrived sooner than anyone else.

They can be the objects of propositional attitudes:

Bill was surprised that John ran.

Finally, this approach accomodates the facts that events can be nominalized and can be referred to pronominally:

> John's running tired him out.
> John ran, and Bill saw it.

But virtually every predication that can be made in natural language can be specified as to time and place, be modified adverbially, function as a cause or effect of something else, be the object of a propositional attitude, be nominalized, and be referred to by a pronoun. It is therefore convenient to extend Davidson's approach to all predications. That is, corresponding to any predication that can be made in natural language, we will say there is an event, or state, or condition, or situation, or "eventuality", or whatever, in the world that it refers to. This approach might be called "ontological promiscuity". One abandons all ontological scruples.

Thus we would like to have in our logical notation the possibility of an extra argument in each predication referring to the "condition" that exists when that predication is true. However, especially for expository convenience, we would like to retain the option of not specifying that extra argument when it is not needed and would only get in our way. Hence, I propose a logical notation that provides two sets of predicates that are systematically related, by introducing what might be called a "nominalization" operator $'$. Corresponding to every n-ary predicate p there will be an $n+1$-ary predicate p' whose first argument can be thought of as the condition that holds when p is true of the subsequent arguments. Thus, if $run(J)$ means that John runs, $run'(E, J)$ means that E is a running event by John, or John's running. If $slippery(F)$ means that floor F is slippery, then $slippery'(E, F)$ means that E is the condition of F's being slippery, or F's slipperiness. The effect of this notational maneuver is to provide handles by which various predications can be grasped by higher predications. A similar approach has been used in many AI systems.

In discourse one not only makes predications about such ephemera as events, states and conditions. One also refers to entities that do not actually exist. Our notation must thus have a way of referring to such entities. We therefore take our model to be a Platonic universe which contains everything that can be spoken of – objects, events, states, conditions – whether they exist in the real world or not. It then may or may not be a property of such entities that they exist in the real world. In the sentence

(1) John worships Zeus,

the worshipping event and John, but not Zeus, exist in the real world, but all three exist in the (overpopulated) Platonic universe. Similarly, in

John wants to fly.

John's flying exists in the Platonic universe but not in the real world.[1,2]

The logical notation then is just first-order predicate calculus, where the universe of discourse is a rich set of individuals, which are real, possible and even impossible objects, events, conditions, eventualities, and so on.

Existence and truth in the actual universe are treated as predications about individuals in the Platonic universe. For this purpose, we use a predicate $Exist$. The formula $Exist(JOHN)$ says that the individual in the Platonic universe denoted by $JOHN$ exists in the actual universe.[3] The formula

(2) $Exist(E) \wedge run'(E, JOHN)$

says that the condition E of John's running exists in the actual universe, or more simply that "John runs" is true, or still more simply, that John runs. A shorter way to write it is $run(JOHN)$.

Although for a simple sentence like "John runs", a logical form like (2) seems a bit overblown, when we come to real sentences in English discourse with their variety of tenses, modalities and adverbial modifiers, the more elaborated logical form is necessary. Adopting the notation of (2) has the effect of splitting a sentence into its propositional content – $run'(E, JOHN)$ – and its assertional claim – $Exist(E)$. This frequently turns out to be useful, as the latter is often in doubt until substantial work has been done by discourse interpretation processes. An entire sentence may be embedded within an indirect proof or other extended counterfactual.

We are now in a position to state formally the systematic relation between the unprimed and primed predicates as an axiom schema. For every n-ary predicate p,

[1] One need not adhere to Platonism to accept the Platonic universe. It can be viewed as a socially constituted, or conventional, construction, which is nevertheless highly constrained by the way the (not directly accessible) material world is. The degree of constraint is variable. We are more constrained by the material world to believe in trees and chairs, less so to believe in patriotism or ghosts.

[2] The reader might choose to think of the Platonic universe as the universe of possible individuals, although I do not want to exclude *logically* impossible individuals, such as the condition John believes to exist when he believes $6 + 7 = 15$.

[3] McCarthy (1977) employs a similar technique.

$$(\forall x_1, ..., x_n)p(x_1, ..., x_n) \supset (\exists e)Exist(e) \wedge p'(e, x_1, ..., x_n)$$

That is, if p is true of $x_1, ..., x_n$, then there is a condition e of p's being true of $x_1, ..., x_n$, and e exists. Conversely,

$$(\forall e, x_1, ..., x_n)Exist(e) \wedge p'(e, x_1, ..., x_n) \supset p(x_1, ..., x_n)$$

That is, if e is the condition of p's being true of $x_1, ..., x_n$, and e exists, then p is true of $x_1, ..., x_n$. We can compress these axiom schemas into one formula:

(3) $(\forall x_1, ..., x_n)p(x_1, ..., x_n) \equiv (\exists e)Exist(e) \wedge p'(e, x_1, ..., x_n)$

A sentence in English asserts the existence of one or more eventualities in the real world, and this may or may not imply the existence of other individuals. The logical form of sentence (1) is

$$Exist(E) \wedge worship'(E, JOHN, ZEUS)$$

This implies $Exist(JOHN)$ but not $Exist(ZEUS)$. Similarly, the logical form of "John wants to fly" is

$$Exist(E_2) \wedge want'(E_2, JOHN, E_1) \wedge fly'(E_1, JOHN)$$

This implies $Exist(JOHN)$ but not $Exist(E_1)$. When the existence of the condition corresponding to some predication implies the existence of one of the arguments of the predication, we will say that the predicate is *transparent* in that argument, and *opaque* otherwise.[4] Thus, *worship* and *want* are transparent in their first arguments and opaque in their second arguments. In general if a predicate p is transparent in its nth argument x, this can be encoded by the axiom

$$(\forall e, ..., x, ...)p'(e, ..., x, ...) \wedge Exist(e) \supset Exist(x)$$[5]

That is, if e is p's being true of x and e exists, then x exists. Equivalently,

$$(\forall ..., x, ...)p(..., x, ...) \supset Exist(x)$$

In the absence of such axioms, predicates are assumed to be opaque.

The following sentence illustrates the extent to which we must have a way of representing existent and nonexistent states and events in ordinary discourse.

[4] More properly, we should say "existentially transparent" and "existentially opaque", since this notion does not coincide exactly with *referential transparency*.

[5] Quantification in this notation is always over entities in the Platonic universe. Existence in the real world is expressed by predicates, in particular the predicate *Exist*.

(4) The government has repeatedly refused to deny that Prime Minister Margaret Thatcher vetoed the Channel Tunnel at her summit meeting with President Mitterand on 18 May, as *New Scientist* revealed last week.[6]

In addition to the ordinary individuals Margaret Thatcher and President Mitterand and the corporate entity *New Scientist*, there are the intervals of time 18 May and "last week", the as yet nonexistent entity, the Channel Tunnel, an individual revealing event and the complex event of the summit meeting, which actually occurred, a set of real refusals distributed across time in a particular way, a denial event which did not occur, and a vetoing event which may or may not have occurred.

Let us take $Past(E_6)$ to mean that E_6 existed in the past and $Perfect(E_1)$ to mean what the perfect tense means, roughly, that E_1 existed in the past and may not yet be completed. The representation of just the verb, nominalizations, adverbials and tenses of sentence (4) is as follows:

$$\begin{aligned} & Perfect(E_1) \wedge repeated(E_1) \wedge refuse'(E_1, GOVT, E_2) \\ & \quad \wedge deny'(E_2, GOVT, E_3) \wedge veto'(E_3, MT, CT) \\ & \quad \wedge at'(E_4, E_3, E_5) \wedge meet'(E_5, MT, PM) \\ & \quad \wedge on(E_5, 18MAY) \wedge Past(E_6) \wedge reveal'(E_6, NS, E_3) \\ & \quad \wedge last\text{-}week(E_6) \end{aligned}$$

Of the various entities referred to, the sentence, via unprimed predicates, asserts the existence of a typical refusal E_1 in a set of refusals and the revelation E_6. The existence of the refusal implies the existence of the government. It does not imply the existence of the denial; quite the opposite. It may suggest the existence of the veto, but certainly does not imply it. The revelation E_6, however, implies the existence of both the *New Scientist* NS and the *at* relation E_4, which in turn implies the existence of the veto and the meeting. These then imply the existence of Margaret Thatcher MT and President Mitterand PM, but not the Channel Tunnel CT. Of course, we know about the existence of some of these entities, such as Margaret Thatcher and President Mitterand, for reasons other than the transparency of predicates.

Sentence (4) shows that virtually anything can be embedded in a higher predication. This is the reason, in the logical notation, for flattening everything into predications about individuals.

[6]This sentence is taken from the *New Scientist*, June 3, 1982 (p. 632). I am indebted to Paul Martin for calling it to my attention.

There are four serious problems that must be dealt with if this approach is to work – quantifiers, opaque adverbials, the distinction between *de re* and *de dicto* readings of belief reports, and the problem of identity in intensional contexts.

I have described a solution to the quantifier problem elsewhere (Hobbs, 1983). Briefly, universally quantified variables are reified as typical elements of sets, existential quantification inside the scope of universally quantified variables are handled by means of dependency functions, and the quantifier structure of sentences is encoded in indices on predicates. In this paper I will address only the other three problems in detail.

3 Opaque Adverbials

It seems reasonably natural to treat transparent adverbials as properties of events. For opaque adverbials, like "almost", it seems less natural, and one is inclined to follow Reichenbach (1947) in treating them as functionals mapping predicates into predicates. Thus,

John is almost a man.

would be represented

$$almost(man)(J)$$

That is, *almost* maps the predicate *man* into the predicate "almost a man", which is then applied to John.

This representation is undesirable for our purposes since it is not first-order. It would be preferable to treat opaque operators as we do transparent ones, as properties of events or conditions. The sentence would be represented

$$almost(E) \wedge man'(E, J)$$

But does this get us into difficulty?

First note that this representation does not imply that John is a man, for we have not asserted E's existence in the real world, and *almost* is opaque and does not imply its argument's existence.

But is there enough information in E to allow one to determine the truth value of $almost(E)$ in isolation, without appeal to other facts? The answer is that there could be. We can construct a model in which for every functional F there is a corresponding equivalent predicate q, such that

$$(\forall p, x)(F(p)(x) \equiv (\exists e)q(e) \wedge p'(e, x))$$

The existence of the model shows that this condition is not necessarily contradictory.

Let the universe of discourse D be the class of finite sets built out of a finite set of urelements. The interpretation of a constant X will be some element of D; call it $I(X)$. The interpretation of a monadic predicate p will a subset of D; call it $I(p)$. Then if E is such that $p'(E, X)$, we define the interpretation of E to be $< I(p), I(X) >$.

Now suppose we have a functional F mapping predicates into predicates. We can define the corresponding predicate q to be such that

> $q(E)$ is true iff there are a predicate p and a constant X where the interpretation of E is $< I(p), I(X) >$ and $F(p)(X)$ is true.

The fact that we can define such a predicate q in a moderately rich model means that we are licensed to treat opaque adverbials as properties of events and conditions.

The purpose of this exercise is only to show the viability of the approach. I am not claiming that a running event *is* an ordered pair of the runner and the set of all runners, although it should be harmless enough for those irredeemably committed to set-theoretic semantics to view it like that.

It should be noted that this treatment of adverbials has consequences for the individuating criteria on eventualities. We can say "John is almost a man" without wishing to imply "John is almost a mammal," so we would not want to say that John's being a man is the *same* condition as his being a mammal. We are forced, though not unwillingly, into a position of individuating eventualities according to very fine-grained criteria.

4 *De Re* and *De Dicto* Belief Reports

The next problem concerns the distinction (due to Quine (1956)) between *de re* and *de dicto* belief reports. A belief report like

(5) John believes a man at the next table is a spy.

has two interpretations. The *de dicto* interpretation is likely in the circumstance in which John and some man are at adjacent tables and John observes suspicious behavior. The *de re* interpretation is likely if some man is sitting at the table next to the speaker of the sentence, and John is nowhere around but knows the man otherwise and suspects him to be a spy. A sentence that very nearly forces the *de re* reading is

John believes Bill's mistress is Bill's wife.[7]

whereas the sentence

John believes Russian consulate employees are spies.

strongly indicates a *de dicto* reading. In the *de re* reading of (5), John is not necessarily taken to know that the man is in fact at the next table, but he is normally assumed to be able to identify the man somehow. More on "identify" below. In the *de dicto* reading John believes there is a man who is both at the next table and a spy, but may be otherwise unable to identify the man. The *de re* reading of (5) is usually taken to support the inference

(6) There is someone John believes to be a spy.

whereas the *de dicto* reading supports the weaker inference

(7) John believes that someone is a spy.

As Quine has pointed out, as usually interpreted, the first of these sentences is false for most of us, the second one true. A common notational maneuver (though one that Quine rejects) is to represent this distinction as a scope ambiguity. Sentence (6) is encoded as (8) and (7) as (9):

(8) $(\exists x) believe(J, spy(x))$

(9) $believe(J, (\exists x) spy(x))$

If one adopts this notation and stipulates what the expressions mean, then there are certainly distinct ways of representing the two sentences. But the interpretation of the two expressions is not obvious. It is not obvious for example that (8) could not cover the case where there is an individual such that John believes him to be a spy but has never seen him and knows absolutely nothing else about him – not his name, nor his appearance, nor his location at any point in time – beyond the fact that he is a spy.

In fact, the notation we propose takes (8) to be the most neutral representation. Since quantification is over entities in the Platonic universe,

[7] This example is due to Moore and Hendrix (1982).

(8) says that there is some entity in the Platonic universe such that John believes of that entity that it is a spy. Expression (8) commits us to no other beliefs on the part of John. When understood in this way, expression (8) is a representation of what is conveyed in a *de dicto* belief report. Translated into the flat notation and introducing a constant for the existentially quantified variable, (8) becomes

(10) $believe(J, P) \wedge spy'(P, S)$

Anything else that John believes about this entity must be stated explicitly. In particular, the *de dicto* reading of (5) would be represented by something like

(11) $believe(J, P) \wedge spy'(P, S) \wedge believe(J, Q) \wedge at'(Q, S, T)$

where T is the next table. That is, John believes that S is a spy and that S is at the next table. John may know many other properties about S and still fall short of knowing *who* the spy is. There is a range of possibilities for John's knowledge, from the bare statements of (10) and (11) that correspond to a *de dicto* reading to the full-blown knowledge of S's identity that is normally present in a *de re* reading. In fact, an FBI agent would progress through just such a range of belief states on his way to identifying the spy.

To state John's knowledge of S's identity properly, we would have to state explicitly John's belief in a potentially very large collection of properties of the spy. To arrive at a succinct way of representing knowledge of identity in our notation, let us consider the two pairs of equivalent sentences:

What is that?
Identify that.

The FBI doesn't know who the spy is.
The FBI doesn't know the spy's identity.

The answer to the question "Who are you?" and what is required before we can say that we know *who* someone is or that we know their identity is a highly context-dependent matter. Several years ago, before I had ever seen Kripke, if someone had asked me whether I knew who Saul Kripke was, I would have said, "Yes. He's the author of *Naming and Necessity*." Then once I was at a workshop which I knew was being attended by Kripke, but I didn't yet know what he looked like. If someone had asked me whether I knew who Kripke was, I would have had to say, "No." The relevant property in that context was not his authorship of some paper, but any property that distinguished him from the others present, such as "the man in the back row holding a cup of coffee".

Knowledge of a person's identity is then a matter of knowing some context-dependent essential property that serves to identify that person for present purposes – that is, a matter of knowing *who* he or she is.

Therefore, we need a kind of place-holder predicate to stand for this essential property, that in any particular context can be specified more precisely. It happens that English has a morpheme that serves just this function – the morpheme "wh". Let us then posit a predicate *wh* that stands for the contextually determined property or conjunction of properties that would count as an identification in that particular context.

The *de re* reading of (5) is generally taken to include John's knowledge of the identity of the alledged spy. Assuming this, a *de re* belief report would be represented as a conjunction of two beliefs, one for the main predication and the other expressing knowledge of the essential property, the what-ness, of the argument of the predication.

$$believe(J, P) \wedge spy'(P, X) \wedge know(J, Q) \wedge wh'(Q, X)$$

That is, John believes S is a spy and John knows who S is.

However, let us probe this distinction just a little more deeply and in particular call into question whether knowledge of identity is really part of the meaning of the sentence in the *de re* reading. The representation of the *de dicto* reading of 5, I have said, is

$$(12)\ believe(J, P) \wedge spy'(P, S) \wedge believe(J, Q) \wedge at'(Q, S, T)$$

Let us represent the *de re* reading as

$$(13a)\quad believe(J, P) \wedge spy'(P, S) \wedge Exist(Q) \wedge at'(Q, S, T)$$

$$(13b)\qquad \wedge\, know(J, R) \wedge wh'(R, S)$$

What is common to (12) and (13) are the conjuncts $believe(J, P)$, $spy'(P, S)$ and $at'(Q, S, T)$. There is a genuine ambiguity as to whether Q exists in the real world (*de re*) or is merely believed by John (*de dicto*). In addition, (13) includes the conjuncts $know(J, R)$ and $wh'(R, S)$ – line (13b).

But are these necessarily part of the *de re* interpretation of sentence 5? The following example casts doubt on this. Suppose the entire Rotary Club is seated at the table next to the speaker of 5, but John doesn't know this. John believes that some member of the Rotary Club is a spy, but has no idea which one. Sentence 5 describes this situation, and only (13a) holds, not (13b) and not (12). Judgments are sometimes uncertain as to whether sentence 5 is appropriate in these circumstances, but it is certain that the sentence

> John believes someone at the next table is a spy.

is appropriate, and that is sufficient for the argument.

It seems then that the conjuncts $know(J, R)$ and $wh'(R, S)$ are not part of what we want in the initial logical form of the sentence,[8] but only a very common conversational implicature. The reason the implicature is very common is that if John doesn't know that the man is at the next table, there must be some other description under which John is familiar with the man. The story I just told provides such a description, but not one sufficient for identifying the man.

This analysis is attractive since it allows us to view the *de re - de dicto* distinction problem as just one instance of a much more general problem, namely, the existential status of the grammatically subordinated material in sentences. Generally, such material takes on the tense of the sentence. Thus, in

> The boy built the boat.

a building event by x of y takes place in the past, and we assume that x was a boy in the past, at the time of the building. But in

> Many rich men studied computer science in college.

the most natural reading is not that the men were rich when they were studying computer science but that they are rich now. In

> The flower is artificial.

there is an entity x which is described as a flower, and x exists, but its "flower-ness" does not exist in the real world. Rather, it is a condition which is embedded in the opaque predicate "artificial".

It was stated above that the representation (10) for the *de dicto* reading conveys no properties of S other than that John believes him to be a spy. In particular, it does not convey S's existence in the real world. S thus refers to a possible individual, who may turn out to be actual if, for example, John ever comes to be able to identify the person whom he believes to be

[8] Another way of putting it: they are not part of the literal meaning of the sentence.

the spy, or if there is some actual spy who has given John good cause for his suspicions.

However, S may not be actual, only possible. Suppose this is the case. One common objection to possible individuals is that they may seem to violate the Law of the Excluded Middle. Is S married or not married? Our intuition is that the question is inappropriate, and indeed the answer given in our formalism has this flavor. By axiom (3), $married(S)$ is really just an abbreviation for $married'(E, S) \wedge Exist(E)$. This is false, for the existence of E in the real world would imply the existence of S. So $married(S)$ is also false. But its falsity has nothing to do with S's marital status, only his existential status. The predication $unmarried(S)$ is false for the same reason. The primed predicates are basic, and for them the problem of the excluded middle does not arise. The predication $married'(E, S)$ is true or false depending on whether E is the condition of S's being married. An unprimed, transparent predicate carries along with it the existence of its arguments, and it can fail to be true of an entity either through the entity being actual but not having that property or through the nonexistence of the entity.

5 Identity in Belief Contexts

The final problem I will consider arises in *de dicto* belief reports. It is the problem of identity in intensional contexts, raised by Frege (1892). One way of stating the problem is this. Why is it that if

(14) John believes the Evening Star is rising.

and if the Evening Star is identical to the Morning Star, it is not necessarily true that

(15) John believes the Morning Star is rising.

By Leibniz's Law, we ought to be able to substitute for an entity any entity that is identical to it.

This puzzle survives translation into the logical notation, if John knows of the existence of the Morning Star and if proper names are unique. The representation for (the *de dicto* reading of) sentence (14) is

(16) $believe(J, P_1) \wedge rise'(P_1, ES) \wedge believe(J, Q_1)$
$\wedge Evening\text{-}Star'(Q_1, ES)$

John's belief in the Morning Star would be represented

$$believe(J, Q_2) \wedge Morning\text{-}Star'(Q_2, MS)$$

The existence of the Evening Star and the Morning Star is expressed by

$$Exist(Q_1) \wedge Exist(Q_2)$$

The uniqueness of the proper name "Evening Star" is expressed by the axiom

$$(\forall x, y) Evening\text{-}Star(x) \wedge Evening\text{-}Star(y) \supset x = y$$

The identity of the Evening Star and the Morning Star is expressed

$$(\forall x) Evening\text{-}Star(x) \equiv Morning\text{-}Star(x)$$

From all of this we can infer that the Morning Star MS is also an Evening Star and hence is identical to ES, and hence can be substituted into $rise'(P_1, ES)$ to give $rise'(P_1, MS)$. Then we have

$$believe(J, P_1) \wedge rise'(P_1, MS) \wedge believe(J, Q_2) \\ \wedge Morning\text{-}Star'(Q_2, MS)$$

This is a representation for the paradoxical sentence (15).

There are three possibilities for dealing with this problem. The first is to discard or restrict Leibniz's Law. The second is to deny that the Evening Star and the Morning Star are identical as entities in the Platonic universe; they only happen to be identical in the real world, and that is not sufficient for intersubstitutivity. The third is to deny that expression (16) represents sentence (14) because "the Evening Star" in (14) does not refer to what it seems to refer to.

The first possibility is the approach of researchers who treat belief as an operator rather than as a predicate, and then restrict substitution inside the operator.[9] We cannot avail ourselves of this solution because of the flatness of our notation. The predicate *rise* is surely referentially transparent, so if ES and MS are identical, MS can be substituted for ES in the expression $rise'(P_1, ES)$ to give $rise'(P_1, MS)$. Then the expression $believe(J, P_1)$ would not even require substitution to be a belief about the Morning Star.

In any case, this approach does not seem wise in view of the central importance played in discourse interpretation by the identity of differently

[9] This is a purely syntactic approach, and when one tries to construct a semantics for it, one is generally driven to the third possibility.

presented entities, i.e. by coreference. Free intersubstitutibility of identicals seems a desirable property to preserve.

The second possible answer to Frege's problem is to say that in the Platonic universe, the Morning Star and the Evening Star are different entities. It just happens that in the real world they are identical. But it is not true that $ES = MS$, for equality, like quantification, is over entities in the Platonic universe. The fact that ES and MS are identical in the real world (call this relation *rw-identical*) must be stated explicitly, say, by the expression

$$rw\text{-}identical(ES, MS)$$

or more properly,

$$(\forall x, y) Morning\text{-}Star(x) \wedge Evening\text{-}Star(y) \\ \quad \supset rw\text{-}identical(x, y)$$

For reasoning about "rw-identical" entities, that is, Platonic entities that are identical in the real world, we may take the following approach. Substitution in referentially transparent contexts would be achieved by use of the axiom schema

$$(17)\ (\forall e_1, e_3, e_4, ...) p'(e_1, ..., e_3, ...) \wedge rw\text{-}identical(e_4, e_3) \\ \quad \supset (\exists e_2) p'(e_2, ..., e_4, ...) \wedge rw\text{-}identical(e_2, e_1)$$

where e_3 is the kth argument of p and p is referentially transparent in its kth argument. That is, if e_1 is p's being true of e_3 and e_3 and e_4 are identical in the real world, then there is a condition e_2 of p's being true of e_4, and e_2 is identical to e_1 in the real world. Substitution of "rw-identicals" in a condition results not in the same condition but in an "rw-identical" condition. There would be such an axiom for the first argument of *believe* but not for its referentially opaque second argument.

Axioms will express the fact that *rw-identical* is an equivalence relation:

$$(\forall x) rw\text{-}identical(x, x) \\ (\forall x, y) rw\text{-}identical(x, y) \supset rw\text{-}identical(y, x) \\ (\forall x, y, z) rw\text{-}identical(x, y) \wedge rw\text{-}identical(y, z) \\ \quad \supset rw\text{-}identical(x, z)$$

Finally, the following axiom, together with axiom (17), would express Leibniz's Law:

$$(\forall e_1, e_2) rw\text{-}identical(e_1, e_2) \supset (Exist(e_1) \equiv Exist(e_2))$$

From all of this we can prove that if the Evening Star rises then the Morning Star rises, but we cannot prove from John's belief that the Evening Star rises that John believes the Morning Star rises. If John knows the Morning Star and the Evening Star are identical, and he knows axiom (17), then his belief that the Morning Star rises can be proved as one would prove his belief in the consequences of any other syllogism whose premises he believed, in accordance with a treatment of reasoning about belief developed in a longer version of this paper.

This solution is in the spirit of our whole representational approach in that it forces us to be painfully explicit about everything. The notation does no magic for us. There is a significant cost associated with this solution, however. When proper names are represented as predicates and not as constants, the natural way to state the uniqueness of proper names is by means of axioms of the following sort:

$$(\forall x, y) \mathit{Evening\text{-}Star}(x) \wedge \mathit{Evening\text{-}Star}(y) \supset x = y$$

But since from the axioms for *rw-identical* we can show that $\mathit{Evening\text{-}Star}(MS)$, it would follow that $MS = ES$. We must thus restate the axiom for the uniqueness of proper names as

$$(\forall x, y) \mathit{Evening\text{-}Star}(x) \wedge \mathit{Evening\text{-}Star}(y) \\ \supset \mathit{rw\text{-}identical}(x, y)$$

A similar modification must be made for functions. Since we are using only predicates, the uniqueness of the value of a function must be encoded with an axiom like

$$(\forall x, y, z) \mathit{father}(x, z) \wedge \mathit{father}(y, z) \supset x = y$$

If x and y are both fathers of z, then x and y are the same. This would have to be replaced by the axiom

$$(\forall x, y, z) \mathit{father}(x, z) \wedge \mathit{father}(y, z) \supset \mathit{rw\text{-}identical}(x, y)$$

The very common problems involving reasoning about equality, which can be done efficiently, are thus translated into problems involving reasoning about the predicate *rw-identical*, which is very cumbersome.

One way to view this second solution is as a fix to the first solution. For "=" we substitute the relation *rw-identical*, and by means of axiom schema (17), we force substitutions to propagate to the eventualities they occur in, and we force the distinction between referentially transparent and referentially opaque predicates to be made explicitly. It is thus an indirect way of rejecting Leibniz' Law.

The third solution is to say that "the Evening Star" in sentence (14) does not really refer to the Evening Star, but to some abstract entity somehow related to the Evening Star. That is, sentence (14) is really an example of metonymy. This may seem counterintuitive, and even bizarre, at first blush. But in fact the most widely accepted classical solutions to the problem of identity are of this flavor. For Frege (1892) "the Evening Star" in sentence (14) does not refer to the Evening Star but to the *sense* of the phrase "the Evening Star". In a more recent approach, Zalta (1983) takes such noun phrases to refer to "abstract objects" related to the real object. In both approaches noun phrases in intensional contexts refer to senses or abstract objects, while other noun phrases refer to actual entities, and so it is necessary to specify which predicates are intensional. In a Montagovian approach, "the Evening Star" would be taken to refer to the *intension* of the Evening Star, not its *extension* in the real world, and noun phrases would *always* be taken to refer to intensions, although for nonintensional predicates there would be meaning postulates that make this equivalent to reference to extensions.

Thus, in all these approaches intensional and extensional predicates must be distinguished explicitly, and noun phrases in intensional contexts are systematically interpreted metonymically.

It would be easy enough in our framework to implement these approaches. We can define a function α of three arguments – the actual entity, the cognizer, and the condition used to describe the entity – that returns the sense, or intension, or abstract entity, corresponding to the actual entity for that cognizer and that condition. Sentence (14) would be represented, not as (16), but as

(18) $believe(J, P_1) \wedge rise'(P_1, \alpha(ES, J, Q_1)) \wedge believe(J, Q_1)$
$\wedge Evening\text{-}Star'(Q_1, ES)$

I tend to prefer to think of the value of $\alpha(ES, J, Q_1)$ as an abstract entity. Whatever it is, it is necessary that the value of $\alpha(ES, J, Q_1)$ be something different from the value of $\alpha(ES, J, Q_2)$ where $Morning\text{-}Star'(Q_2, ES)$. That is, different abstract objects must correspond to the condition Q_1 of being the Evening Star and the condition Q_2 of being the Morning Star. It is because of this feature that we escape the problem of intersubstitutivity of identicals, for substitution of MS for ES in (18) yields "... $\wedge$

$rise'(P_1, \alpha(MS, J, Q_1)) \wedge ...$" rather than "$... \wedge rise'(P_1, \alpha(MS, J, Q_2)) \wedge ...$", which would be the representation of sentence (15).

The difficulty with this approach is that it makes the interpretation of noun phrases dependent on their embedding context:

Intensional context $\rightarrow$ metonymic interpretation
Extensional context $\rightarrow$ nonmetonymic interpretation

It thus violates, though not seriously, the naive compositionality that I have been at so many pains to preserve. Metonymy is a very common phenomenon in discourse, but I prefer to think of it as occurring irregularly, and not as signalled systematically by other elements in the sentence.

Having laid out the three possible solutions and their shortcomings, I find that I would like to avoid the problem of identity altogether. The third approach suggests a ruse for doing so. We can assume that, in general, (16) is the representation of sentence (14). We invoke no extra complications where we don't have to. When, in interpreting the text, we encounter a difficulty resulting from the problem of identity, we can go back and revise our interpretation of (14), by assuming the reference must have been a metonymic one to the abstract entity and not to the actual entity. In those cases it would be as if we are saying, "John couldn't believe about the Evening Star itself that it is rising. The paradox shows that he is insufficiently acquainted with the Evening Star to refer to it directly. He must be talking about an abstract entity related to the Evening Star." My guess is that we will not have to resort to this ruse often, for I suspect the problem rarely arises in actual discourse interpretation.

6 The Role of Semantics

Let me close by making some comments about ways of doing semantics. Semantics is the attempted specification of the relation between language and the world. However, this requires a theory of the world. There is a spectrum of choices one can make in this regard. At one end of the spectrum – let's say the right end – one can adopt the "correct" theory of the world, the theory given by quantum mechanics and the other sciences. If one does this, semantics becomes impossible because it is no less than all of science, a fact that has led Fodor (1980) to express some despair. There's too much of a mismatch between the way we view the world and the way the world really is. At the left end, one can assume a theory of the world that is isomorphic to the way we talk about it. What I have been doing in this paper, in fact, is an effort to work out the details in such a theory. In this case, semantics becomes very nearly trivial. Most activity in semantics today is slightly to the right of the extreme left end of this spectrum. One makes certain assumptions about the nature of the world that closely reflect language, and doesn't make certain other assumptions. Where one has failed to make the necessary assumptions, puzzles appear, and semantics becomes an effort to solve those puzzles. Nevertheless, it fails to move far enough away from language to represent significant progress toward the right end of the spectrum. The position I advocate is that there is no reason to make our task more difficult. We will have puzzles enough to solve when we get to discourse.

Acknowledgements

I have profited from discussions about this work with Chris Menzel, Bob Moore, Stan Rosenschein, and Ed Zalta. This research was supported by NIH Grant LM03611 from the National Library of Medicine, by Grant IST-8209346 from the National Science Foundation, and by a gift from the Systems Development Foundation.

References

[1] Bobrow, Daniel G. and Terry Winograd, 1977. "An Overview of KRL, A Knowledge Representation Language", *Cognitive Science*, vol. 1, pp. 3-46.

[2] Davidson, Donald, 1967. "The Logical Form of Action Sentences", in N. Rescher, ed., *The Logic of Decision and Action*, pp. 81-95, University of Pittsburgh Press, Pittsburgh, Pennsylvania.

[3] Fodor, J. A., 1980. "Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology", *The Behavioral and Brain Sciences*, vol. 3, no. 1, March, 1980.

[4] Frege, Gotlieb, 1892. "On Sense and Nominatum", in H. Feigel and W. Sellars, ed., *Readings in Philosophical Analysis*, pp. 85-102, Appleton-Century-Croft, Inc., New York, 1949.

[5] Hayes, Patrick J., 1979. "The Logic of Frames", in D. Metzing, ed., *Frame Conceptions and Text Understanding*, pp. 46-61, Walter de Gruyter and Company.

[6] Hendrix, Gary G., 1975. "Extending the Utility of Semantic Networks Through Partitioning", *Advance Papers, International Joint Conference on Artificial Intelligence*, Tbilisi, Georgian SSR, pp. 115-121, September, 1975.

[7] Hobbs, Jerry R., 1983. "An Improper Treatment of Quantification in Ordinary English", *Proceedings of the 21st Annual Meeting, Association for Computational Linguistics*, pp. 57-63. Cambridge, Massachusetts, June, 1983.

[8] McCarthy, John, 1977. "Epistemological Problems of Artificial Intelligence", *Proceedings, International Joint Conference on Artificial Intelligence*, pp. 1038-1044, Cambridge, Massachusetts, August, 1977.

[9] Minsky, Marvin, 1975. "A Framework for Representing Knowledge", in Patrick H. Winston, ed., *The Psychology of Computer Vision*, pp. 211-277, McGraw-Hill.

[10] Moore, Robert C., 1980. "Reasoning about Knowledge and Action", SRI International Technical Report 191, October, 1980.

[11] Moore, Robert C. and Gary G. Hendrix, 1982. "Computational Models of Belief and the Semantics of Belief Sentences", in S. Peters and E. Saarinen, eds., *Processes, Beliefs, and Questions*, pp. 107-127, D. Reidel Publishing Company.

[12] Quillian, M. Ross, 1968. "Semantic Memory", in Marvin Minsky, ed., *Semantic Information Processing*, pp. 227-270, MIT Press, Cambridge, Massachusetts.

[13] Quine, Willard V., 1953. "On What There Is", in *From a Logical Point of View*, pp. 1-19, Harvard University Press, Cambridge, Massachusetts.

[14] Quine, Willard V., 1956. "Quantifiers and Propositional Attitudes", *Journal of Philosophy*, vol. 53.

[15] Reichenbach, Hans, 1947. *Elements of Symbolic Logic*, The MacMillan Company.

[16] Schmolze, J. G. and R. J. Brachman, 1982. "Summary of the KL-ONE Language", in *Proceedings, 1981 KL-ONE Workshop*, pp. 231-257, Fairchild Laboratory for Artificial Intelligence Research, Palo Alto, California.

[17] Simmons, Robert F., 1973. "Semantic Networks: Their Computation and Use for Understanding English Sentences", in Roger Schank and Kenneth Colby, eds., *Computer Models of Thought and Language*, pp. 63-113, W. H. Freeman: San Francisco.

[18] Zalta, Edward N., 1983. *Abstract Objects: An Introduction to Axiomatic Metaphysics*, D. Reidel Publishing Company: Dordrecht, Netherlands.