Indexical Expressions in the Scope of Attitude Verbs

Andrew R. Haas*
The University at Albany

A sentential theory of attitudes holds that propositions (the things that agents believe and know) are sentences of a representation language. Given such a theory, it is natural to suggest that the proposition expressed by an utterance of natural language is also a sentence of a representation language. This leads to a straightforward account of the semantics of attitude verbs. However, this kind of theory encounters problems in dealing with indexicals: expressions such as "I," "here," and "now." It is hard to explain how an indexical in the scope of an attitude verb can be opaque. This paper suggests that while the propositions that agents believe and know are sentences, the propositions expressed by utterances are not sentences: they are singular propositions, of the type used in Kaplan's theory of direct reference. Drawing on recent work in situation semantics, this paper shows how such a theory can describe the semantics of attitude verbs and account for the opacity of indexicals in the scope of these verbs.

1. Introduction

A sentential theory of attitudes holds that propositions (the things that agents believe and know) are sentences of a representation language. The idea has an obvious appeal to AI workers, who rely heavily on representation languages. Moore and Hendrix (1982) gave the classic statement of the case for a sentential theory of attitudes. Konolige (1986), Haas (1986), and Perlis (1988) are among the authors who have applied sentential theories to AI problems.

A central problem for any theory of attitudes is the **opacity** of NPs in the scope of attitude verbs. Suppose we take a sentence containing an occurrence of the noun phrase NP_1 and we replace that occurrence with another noun phrase NP_2 that refers to the same individual. In many cases, the substitution will preserve the truth value of the sentence. If "Superman" and "Clark Kent" refer to the same man, then "Superman was born on Krypton" and "Clark Kent was born on Krypton" have the same truth value. However, if NP_1 and NP_2 occur in the scope of an attitude verb, the two sentences might have different truth values. If "Superman" and "Clark Kent" refer to the same man, and Lois Lane knows this man, the sentence "Lois thinks that Superman was born on Krypton" might be true even though "Lois thinks that Clark Kent was born on Krypton" is false. Then the occurrence of "Superman" in the first sentence is called an opaque occurrence.

Current versions of the sentential theory can explain many examples of opacity, but they have trouble when the opaque NP is an **indexical**. "I" and "now" are typical indexicals. When "I" occurs in an English utterance, it refers to the speaker of that utterance. When "now" occurs in an English utterance, it refers to the time when the

* Department of Computer Science, The University at Albany, 1400 Washington Avenue, Albany, NY 12222

speaker produces the utterance. In general, indexicals are words and phrases whose reference is fixed by a property of the utterance they occur in. Demonstratives are a class of phrases akin to indexicals, whose reference often depends on gestures that accompany an utterance. If I point at a book and say "Give me that book," the phrase "that book" refers to the book I pointed at. Indexicals and demonstratives occur constantly in human speech, and they present crucial problems for any theory of propositional attitudes. In particular, it is essential to explain how an indexical in the scope of an attitude verb can be opaque.

If we adopt a sentential theory of attitudes, it is natural to suggest that the proposition expressed by an utterance is also a sentence of a representation language. This leads to a straightforward account of the semantics of attitude verbs. In Section 2 I will describe this approach and show why it fails to explain the opacity of indexicals in the scope of attitude verbs. In Section 3 I will describe a nonsentential semantics for attitude verbs due to Crimmins and Perry (1989). This theory can explain the opacity of indexicals. In Section 4 I will describe a new version of the sentential semantics for attitude verbs, which incorporates some of Crimmins and Perry's ideas. This theory denies that the proposition expressed by an utterance is a sentence in a representation language. It follows (given a sentential theory of attitudes) that an utterance expresses a proposition that no one can believe or know. This sounds paradoxical, but I will argue that the new theory explains the opacity of indexicals while maintaining the advantages of a sentential theory of attitudes.

2. Sentential Semantics for Attitude Verbs—Version One

2.1 Using Quotation to Represent Attitudes

Computational linguists often translate sentences of natural language into sentences of a representation language. This practice suggests a theory: that the proposition expressed by an utterance is a sentence of a representation language. This leads to a straightforward analysis of the semantics of attitude verbs. We introduce the quotation mark '. If Q is a wff or term of our language, then 'Q is a constant that denotes the expression Q. If a sentence P expresses the proposition Q, then an utterance of "Mary believes that P" expresses the proposition bel(mary,'Q). The symbol "bel" is an ordinary predicate letter, not a special modal operator. The first argument of "bel" denotes an agent, and the second argument denotes a sentence. The argument of the quotation mark may itself contain quotation marks: we might represent "John thinks Mary thinks the world is flat" as

1 think(john, 'think(mary 'flat(world))).

This analysis of attitude verbs can certainly account for some examples of opacity. Under this analysis, the sentence "Lois thinks Superman was born on Krypton" might express the proposition

2 think(lois, 'born(superman,krypton))

while "Lois thinks Clark Kent was born on Krypton" might express

3 think(lois, 'born(clark_kent,krypton)).

The terms 'born(superman,krypton) and 'born(clark_kent,krypton) denote two different sentences, so it is quite possible that Lois has one sentence in her knowledge base and not the other.

It is straightforward to introduce this kind of quotation into a first-order language. Suppose we are given a first-order language L_0. We define the first-order languages

$L_1, L_2, \ldots$ by induction as follows: the symbols of L_{i+1} are all symbols of L_i, together with a new constant $'Q$ for each term or wff Q in L_i. Let L be the union of $L_0, L_1, L_2, \ldots$. Then L is a first-order language, and if Q is a term or wff of L, $'Q$ is a constant of L. Let **M** be a structure for L, and suppose that for each term or wff Q in L, the denotation of the constant $'Q$ in **M** is Q. Then we say that **M** is a **quotation structure** for **M**. These definitions do not take us outside of ordinary first-order logic—we have simply defined an interesting subset of the class of all first-order structures.

Given a system of quotation, we can introduce a predicate called "bel" with three arguments: an agent, a sentence, and a time (I omit the time argument in this paper). If this predicate represents belief, we must introduce axioms that allow us to make the desired inferences about beliefs. Perception can create beliefs; inference creates new beliefs from old ones; an agent uses beliefs to choose actions; and so on. Various writers have developed theories of this kind, including the present author (Haas 1986). This paper is about one problem, opacity, and my arguments do not depend on the details of the theory of belief and other attitudes. Therefore I will not develop the theory here.

One issue in the theory of attitudes is so notorious that I must mention it: the paradoxes of self-reference. Quotation by itself creates no paradoxes, but if we introduce a predicate that represents truth or knowledge, it is easy to write plausible-looking axioms that are inconsistent. The literature on this topic is large; see Montague (1974), Kripke (1975), and Barwise and Etchemendy (1987). It was once thought that self-reference paradoxes are particularly troublesome for sentential theories, but des Rivières and Levesque (1986) have corrected this error. Perlis (1988) has shown how we can take a language with quotation and consistently introduce a predicate with some (not all) of the properties we expect of a truth predicate. These issues are important, but they are not closely related to the problem of indexicals and opacity, so I will not consider them further here. See Haas (1991) for more references and discussion.

2.2 Indexicals as Descriptions

The reference of an indexical depends on the utterance it occurs in. Therefore if we intend to study indexicals, we can no longer speak of a sentence expressing a proposition. An utterance expresses a proposition, and different utterances of the same sentence can express different propositions. We must now reformulate the analysis of attitude verbs suggested above. Suppose P is a sentence, and consider the sentence "Mary believes that P." If a speaker utters "Mary believes that P," he or she also utters P. If the utterance of P expresses a proposition Q, the utterance of "Mary believes that P" expresses the proposition bel(mary, $'Q$).

If we translate English sentences into a representation language, we must translate indexicals into expressions of the representation language. One way to do this is to assign a constant to denote each utterance. Suppose the words of a certain utterance are "I am now in London," and the constant that denotes the utterance is U729. Then the proposition the utterance expresses might be

4 $(\exists x, y.\ \text{speaker}(x, \text{U729}) \wedge \text{time}(y, \text{U729}) \wedge \text{in}(x, \text{london}, y))$.

This sentence says that the speaker of U729 is in London at the time when U729 occurs. The wff speaker(x,U729) is a description of the speaker of U729, and the wff time(y,U729) is a description of the time when U729 occurs. This resembles Reichenbach's analysis of indexicals (Reichenbach 1947).

If indexicals translate into descriptions, it is possible to construct an opaque reading for an indexical in the scope of an attitude verb. As an example, suppose Clark

Kent thinks that Lois Lane has discovered his secret identity, but he is mistaken. Kent reports his belief to Jimmy Olsen by pointing at a picture of Superman in costume and saying

5 Lois believes I am that man.

This utterance is false. On the other hand, suppose Kent later says

6 Lois believes I am me.

This utterance is true. The NP "that man" in the first utterance and the NP "me" in the second utterance refer to the same individual, but one utterance is false and the other is true. Therefore the NPs are opaque.

If u is an utterance, let demonstrate(x, u) mean that x is the object that the speaker of u demonstrates to the addressee while uttering u. To get opaque readings for the NPs "that man" and "me" in these utterances, we can assign the following representations. Let the constant U891 denote the first utterance, and let the representation of this utterance be

7 bel(lois, $'(\exists x, y.\ \text{speaker}(x, \text{U891}) \wedge \text{demonstrate}(y, \text{U891}) \wedge x = y))$.

Let the constant U144 denote the second utterance, and let the representation of this utterance be

8 bel(lois, $'(\exists x, y.\ \text{speaker}(x, \text{U144}) \wedge \text{speaker}(y, \text{U144}) \wedge x = y))$.

These two representations describe two quite different sentences. It is quite possible that Lois believes one sentence and not the other, so it is quite possible that (7) is false and (8) is true. So we have opaque readings for the NPs "that man" and "me."

We get the opaque readings for (5) and (6) by putting the descriptions that represent the NPs in the scope of the quotation mark on the second argument of "bel." By placing a description in the scope of the quotation mark, we are claiming that this description occurs in a sentence that Lois believes. For example, (7) entails that Lois believes the sentence

9 $(\exists x, y.\ \text{speaker}(x, \text{U891}) \wedge \text{demonstrate}(y, \text{U891}) \wedge x = y))$

which means that Lois knows (or can easily infer) that the speaker of U891 demonstrated some object to his or her hearer while uttering U891. So if (7) represents the intended meaning of Kent's utterance (5), Kent was implying that Lois knew that he demonstrated some object while producing this utterance. This is clearly wrong. Kent's statement might easily be true even if Lois was not present when Kent uttered U891, and has no idea that the utterance even occurred, let alone that Kent demonstrated anything as he spoke. So (7) cannot be a correct representation of the intended meaning of Kent's utterance.

There is another possible representation of Kent's utterance, one that avoids the false implication of (7). Let subst$(R, [v_1, \ldots, v_k], [u_1, \ldots, u_k])$ be the wff formed by simultaneously substituting the terms $u_1, \ldots, u_k$ for free occurrences of the variables $v_1, \ldots, v_k$ in R. For example, we have subst$('p(x,y), ['x], ['c]) = 'p(c, y)$. Using this notation, we can form a new representation for "Lois believes I am that man"—one that places the description "demonstrate(x, U891)" outside the scope of the quotation mark. The new representation says that there exists a term t_1 of the representation language that denotes the demonstratum of U891, and another term t_2 which denotes the speaker of U891, and Lois believes the sentence $t_1 = t_2$. Following Kaplan (1975), we want to indicate that t_1 and t_2 are not just any terms—they must have some significance. We write name(a, t, x) to mean that t is a term of our representation language, the denotation of t is x, and t contains information that is significant to the agent a. For the moment let us not inquire what this significance amounts to—the question does not effect the present argument.

Using the substitution operator and the predicate "name," we can build the following representation for Kent's utterance (5):

10

$$(\exists x, y, t_1, t_2.\ \text{speaker}(x, \text{U891}) \wedge \text{demonstrate}(y, \text{U891}) \\ \wedge\ \text{name}(\text{lois}, t_1, x) \wedge \text{name}(\text{lois}, t_2, y) \\ \wedge\ \text{bel}(\text{lois}, \text{subst}(`(x = y), [`x, `y], [t_1, t_2])))$$

And for (6):

11

$$(\exists x, y, t_1, t_2.\ \text{speaker}(x, \text{U144}) \wedge \text{speaker}(y, \text{U144}) \\ \wedge\ \text{name}(\text{lois}, t_1, x) \wedge \text{name}(\text{lois}, t_2, y) \\ \wedge\ \text{bel}(\text{lois}, \text{subst}(`(x = y), [`x, `y], [t_1, t_2])))$$

To see that (10) is a plausible representation of utterance (5), suppose that k and s are terms that denote Clark Kent (and therefore denote Superman), and both terms contain information that is significant to Lois. k occurs in the following beliefs of Lois (among others):

12 personal_name(k, "Clark Kent")
13 profession(k, reporter)

while s occurs in these beliefs:

14 personal_name(s, "Superman")
15 profession(s, superhero).

Then name(lois, 'k, k) and name(lois, 's, k) are true. Suppose also (contrary to our story) that Lois believes the sentence k = s. This is precisely the kind of situation that Kent has in mind when he says to himself "Lois has discovered my secret identity," and it is this belief that prompts him to produce utterance (5). Let s be an assignment such that $s(t_1)$ is the constant k and $s(t_2)$ is the constant s. Then the value of the term $\text{subst}(`(x = y), [`x, `y], [t_1, t_2])$ in s is the sentence k = s. Therefore (10) would be true if Lois believed the sentence k = s. In (10) and (11), the translations of the indexicals are outside the scope of the quotation marks. Therefore (10) and (11) do not imply that Lois has any knowledge of the utterances (5) and (6). This kind of representation for attitude reports comes from Kaplan (1975). Haas (1991) presented a formal grammar that builds such representations, using a different mechanism for quotation.

Unfortunately, the representations (10) and (11) do not account for the opacity of the NPs. If "that man" in U891 and "me" in U144 refer to the same individual, then the demonstratum of U891 and the speaker of U144 are the same man. Therefore the wffs demonstrate(y, U891) and speaker(y, U144) have the same denotation: a function that maps a substitution s to True if $s(y)$ is the speaker of U891 (= the demonstratum of U144). The wffs speaker(x, U891) and speaker(x, U144) also have the same denotation. It follows that (10) and (11) must have the same truth value. According to our story, (10) is true—as the reader can see by considering an assignment s such that $s(x) =$ Kent = Superman, $s(y) =$ Kent = Superman, and $s(t_1) = s(t_2) =$ the constant k. Therefore (10) cannot express the intended meaning of Kent's utterance (5)—because the proposition Kent was trying to express is false.

Our attempt to translate indexicals into a representation language has ended in a dilemma. If we represent an indexical as a description, that description must specify the utterance that the indexical occurs in—because the reference of the indexical depends on that utterance. We might represent "I am that man" as "the speaker of U891 is the man demonstrated by the speaker of U891." Suppose the indexical occurs in the scope of an attitude verb, and it is opaque—as in "Lois thinks I am that man." In order to account for the opacity, we must put the description that represents the indexical within the scope of a quotation mark. This entails that the description is part of Lois's belief—which means that Lois knows about the utterance U891. This is clearly wrong. If "Lois thinks I am that man" is true, it doesn't follow that Lois thinks that the speaker of U891 is the same as the individual demonstrated by the speaker of U891. It is quite possible that Lois knows nothing about the utterance U891.

We can get rid of this implication by moving the description up, out of the scope of the quotation mark—but then we lose our explanation of opacity. Either way, we get it wrong. It seems that translating indexicals into descriptions will not help us to explain the opacity of indexicals in the scope of attitude verbs.

3. Direct Reference and Singular Propositions

3.1 Direct Reference and Semantic Interpretation

We have considered a theory claiming that indexicals are represented by descriptions. Let us now consider a very different approach: Kaplan's theory of direct reference (Kaplan 1989). Suppose an utterance contains an indexical expression referring to an entity x. According to Kaplan, this utterance expresses a proposition that contains x itself—not a concept of x, or a description of x in a representation language. Kaplan called such a proposition a **singular** proposition. He writes:

> Don't think of propositions as sets of possible worlds, but rather as structured entities looking something like the sentences which express them. For each occurrence of a singular term in a sentence there will be a corresponding constituent in the proposition expressed....in the case of a singular term which is directly referential, the constituent of the proposition is just the object itself. (Kaplan 1989, p. 494)

Kaplan never denied that a proposition might contain concepts or descriptions—he only said that indexicals refer to entities without introducing any concept or description into the proposition. He argued that the same is true for proper names, but for convenience I will ignore this and translate proper names as constants of the representation language.

Suppose direct reference is right. Consider an utterance that contains indexicals referring to people, places, or times. These people, places, or times must be constituents of the proposition expressed by that utterance. It follows that this proposition is not a sentence of a representation language—because people, places, and times cannot be constituents of sentences. If we maintain a sentential theory, we must conclude that this utterance expresses a proposition that no one can believe or know—because the propositions that people believe and know are sentences, but the proposition the utterance expresses is not a sentence. Then we cannot interpret "Mary believes that P" as meaning that Mary believes the proposition that P expresses. Instead she must believe a sentence that is somehow related to the proposition P expresses. What can this relation be?

3.2 A New Semantics for Attitude Reports

Crimmins and Perry (1989) proposed a new approach to NPs in the scope of attitude reports. They proposed the idea as a modification of situation semantics, but it is easily incorporated into a sentential theory. I will first consider their work and then describe a sentential version of their ideas.

Crimmins and Perry begin by distinguishing between beliefs and propositions. Beliefs are mental entities—they would not exist if there were no thinking beings. A proposition is not a mental entity. It consists of a property or relation and a sequence of arguments. The proposition that John loves Mary consists of the relation of loving and the sequence of John and Mary. This proposition contains the people themselves, not concepts or representations. So it is a singular proposition. It also contains the relation of loving, which is not a symbol or a concept, any more than John or Mary is.

The beliefs that Crimmins and Perry consider in their paper are similar to atomic sentences of a representation language. Each belief involves a k-ary idea (similar to a k-ary predicate of representation language) and a sequence of k notions (similar to closed terms of representation language). To represent the structure of the belief we write

16 $\langle \text{Idea}^k, \text{Notion}_1, \ldots \text{Notion}_k \rangle$.

An idea is normally an idea *of* some property or relation. We write Of(Idea^k) for the property or relation that Idea^k is an idea of. Likewise a notion is usually a notion of something, and we write Of(Notion_1) for the individual that Notion_1 is a notion of. In Crimmins and Perry the object of a notion can change over time, and the function "Of" takes a time as one of its arguments. I have simplified by omitting this argument.

The content of a belief is a proposition. In particular, if (16) describes the structure of a certain belief, then its content is a proposition consisting of the relation Of(Idea^k) and the sequence of individuals Of(Notion_1),. . . ,Of(Notion_k). We write this proposition as

17 $\langle\langle \text{Of}(\text{Idea}^k); \text{Of}(\text{Notion}_1), \ldots, \text{Of}(\text{Notion}_k)\rangle\rangle$.

In this case we say that the notion Notion_j is *responsible* for filling the j-th argument position in proposition (17).

According to Crimmins and Perry, an attitude is a ternary relation. Its arguments are an agent, a proposition, and a sequence of notions. Let us take belief as an example. The belief relation holds between agent a, proposition p, and notions $\text{Notion}_1, \ldots \text{Notion}_k$ iff a has a belief whose content is p, and that belief has the form

18 $\langle \text{Idea}_1, \text{Notion}_1, \ldots \text{Notion}_k \rangle$.

That is, the notions $\text{Notion}_1, \ldots \text{Notion}_k$ are the ones responsible for filling the argument positions in the proposition p. The belief report says that the agent has a belief whose content is p, but it says more: it specifies the notions that appear in that belief.

Consider the Superman example above. Lois has two notions of Superman/Kent—call them k and s. The notion k occurs in the beliefs that Lois would express in English as "Kent was born on Earth" and "Kent cannot fly." The notion s occurs in beliefs that Lois would express as "Superman was born on Krypton" and "Superman can fly." Let "Identity" be the identity relation, which holds between x and y iff $x = y$. When Kent points at the picture and says "Lois thinks I am that man," he asserts that the belief relation holds between Lois, the proposition

19 $\langle\langle \text{Identity}; \text{Superman}, \text{Superman}\rangle\rangle$,

and the sequence of k and s. The assertion is true if Lois has a belief whose content is (19), and that belief has the form

20 $\langle \text{Idea}_1;\text{k,s} \rangle$.

Since the content of this belief is (19), Idea_1 must be Lois's idea of the identity relation. Kent is not just saying that Lois has a belief whose content is (19)—he is also asserting that this belief contains particular notions k and s. Because k and s are distinct, belief (20) is not a trivial inference from the belief that everything is equal to itself.

When Kent says "Lois thinks I am me," he is describing a trivial belief that Lois would express in English as "Kent is Kent." He is asserting that the belief relation holds between Lois, the proposition (19), and the sequence of k and k. The assertion is true if Lois has a belief whose content is (19), and that belief has the form

21 $\langle \text{Idea}_1,\text{k,k} \rangle$.

Then Kent's two assertions express two different propositions. Each proposition consists of the belief relation and a triple containing Lois, proposition (19), and a sequence of two notions. In the first assertion the sequence is [s,k], while in the second assertion it is [k,k]. Since the two assertions express different propositions, it is quite possible that one is true and the other false. Thus Crimmons and Perry are able to explain the opacity of indexicals.

This simple solution to the problem has one drawback. We claim that when Kent points and says "Lois thinks I am that man," he expresses a proposition that refers directly to two notions k and s that appear in Lois's beliefs. The sentence "Lois thinks I am that man" contains a proper name that refers to Lois, and indexicals that refer to Kent, but it contains no word or phrase that refers to k or s. k and s are constituents of the proposition Kent expressed, but his words do not name or describe them in any way. k and s are what Crimmins and Perry call **unarticulated constituents**.

There is clear evidence for unarticulated constituents in cases that do not involve attitudes. For example, the verb "rain" takes two logical arguments: a time and a place. Sometimes they are both articulated: "It rained in New York last night." Often the place is unarticulated. If you return from Bermuda I might meet you and say "Hi, how was your trip?" You might reply "It rained every day." You have expressed a proposition with Bermuda among its constituents, though you did not use any name, description, or indexical expression that refers to Bermuda. Such examples show that we cannot object to the ideas of Crimmins and Perry because they rely on unarticulated constituents.

The reader should note that we must generalize Crimmins and Perry's idea in order to handle quantification into the scope of attitudes. Suppose we have "Every man thinks his wife is smart." Each man has a notion of his wife, and the speaker is presumably referring to a function that maps each man to the correct notion. We cannot represent this example in Crimmins and Perry's notation, because it is not clear how they intend to represent quantifiers.

4. Sentential Semantics for Attitudes—Version Two

4.1 Singular Propositions and Sentences

Crimmins and Perry firmly deny that a proposition is a sentence or an internal representation. Yet in order to account for opacity, they need to use notions, which are internal representations. Since internal representations play an essential role in their explanation of opacity, it is natural to ask if we can find a similar explanation within a sentential theory.

If direct reference is right, the proposition that an utterance expresses cannot be a sentence. However, it might still be a lot like a sentence. Let us assume that a singular proposition is an ordered pair, containing a wff with free variables and an assignment of values to those variables. For example, an utterance of "I am now in London" might express a proposition consisting of the wff in(x,london,y) and a function that maps x to the speaker and y to the time of the utterance. I will write this singular proposition as ('in(x,london,y),f), where f is a function mapping x to the speaker and y to the time. This account of singular propositions is consistent with Kaplan's description: the referents of indexicals are constituents of the proposition, and propositions are "structured entities looking something like the sentences which express them." Kaplan's description obviously gives us a lot of leeway; I am taking advantage of that leeway to make singular propositions as much like sentences as possible. I will show that by adopting this view of singular propositions, we can account for the opacity of indexicals within a sentential theory of attitudes.

Suppose, as in Section 2.2 above, that Lois's beliefs contain two constants k and s that denote Superman/Kent, and these constants occur in beliefs (12–15), among others.

12 personal_name(k, "Clark Kent")
13 profession(k,reporter)
14 personal_name(s,"Superman")
15 profession(s,superhero)

When Kent says "Lois thinks I am that man," he is expressing a singular proposition that refers directly to the terms k and s. This proposition says that k and s denote Superman/Kent, and that Lois believes the sentence k = s. The proposition is (P,f), where P is the wff

22 $\text{denote}(z,x) \wedge \text{denote}(w,y) \wedge \text{believe(lois,subst(}`(x=y),[`x,`y],[z,w]))$

and f is the function that maps x and y to Superman, z to k, and w to s. Note that the variables x and y have both quoted and unquoted occurrences in this wff. The unquoted occurrences refer directly to Superman, while the quoted occurrences are in the arguments of the function letter "subst." These occurrences serve to indicate the places where the ground terms z and w, denoting Superman, occur in Lois's belief. This double use of the variables is a little confusing, but it will simplify the semantic interpretation rules for attitude verbs.

When Kent makes the uninteresting assertion "Lois thinks I am me," he expresses a different singular proposition. This proposition is (P,g), where P is wff (22) and g is a function that maps x and y to Kent, z to the constant k, and w also to the constant k. These two singular propositions contain the same wff—they differ only in the values they assign to the variables in that wff. The second proposition says that Lois believes the trivial sentence k = k. Since they describe different beliefs, it is quite possible that one is true and the other false—even if Kent and Superman are the same person. So this analysis can account for opaque indexicals.

Suppose we accept direct reference and also maintain a sentential theory of attitudes. Then in "Mary believes that P," the utterance of P expresses a singular proposition, while Mary's belief is a sentence. We must somehow construct this sentence from the singular proposition expressed by P. Our proposal for this construction is in two parts. First, we suggest that a singular proposition is a pair (Q,f), where Q is a wff and f is a function that maps the free variables of Q to their values. This is consistent with Kaplan's ideas about singular propositions, and it suggests an obvious way of constructing a sentence: we replace each free variable x in Q with a ground term that denotes the value $f(x)$. To complete the construction, we must choose a ground term for each free variable of Q. Here we rely on Crimmins and Perry's notion of unarticulated constituents. We claim that the ground terms we need are constituents of the proposition expressed by "Mary believes that P"—even though they are not named or described by any words of the utterance. Given these ground terms, we take the proposition expressed by P and construct the sentence that Mary believes. In this way we can reconcile direct reference with a sentential theory of attitudes.

4.2 Further Examples

Suppose John says "I am smart." Hearing his words, one would naturally describe John's belief by saying "He thinks he is smart." In this sentence the pronoun "he" appears in the scope of an attitude verb, and it represents the subject's use of a first person pronoun. Castañeda (1968) coined the term **quasi-indicator** for an occurrence of a pronoun that is in the scope of an attitude verb and that represents the subject's use of some indexical expression.

We can analyze this example as follows. John's belief is the sentence smart(i), where i is the constant that John normally uses to refer to himself. This constant will play a special role in perception, introspection, and planning. If John sees a lion in front of him, his sensors will create the belief $(\exists x.\ \mathrm{lion}(x) \land \mathrm{in\_front\_of}(x, \mathrm{i}))$. If he learns by introspection that he believes the world is round, he will form the belief bel(i, 'round(world)). If John wants to achieve a goal represented by a sentence P, he will try to identify an action x such that $(\mathrm{do}(\mathrm{i}, x) \rightarrow P)$ is true. See Haas (1986) for further examples of how an agent can use a standard constant to refer to himself or herself. Such a constant is called the agent's **selfname**.

An utterance of "John thinks he is smart" would normally express a singular proposition (Q, f), where Q is the wff

23 denote(z,john) $\land$ believe(john,subst('smart(x),['x], [z]))

and f is a function that maps the variable z to John's selfname. It is possible that there are other terms in John's internal language that denote John, and in some unusual situation, an utterance of "John thinks he is smart" might refer to one of those other terms. But a person who says "John thinks he is smart" would normally be referring to John's selfname.

Let us give an informal statement of the semantic rule for attitude verbs that we are proposing. Consider a sentence of the form "John believes that P," where P is a clause. P will express a singular proposition (Q, g), where Q is a wff with free variables $x_1, \ldots, x_n$ and g is a function mapping these variables to their values. We can refer to Q by the constant 'Q. Let $t_1, \ldots, t_n$ be closed terms of representation language—terms that (in the speaker's opinion) appear in John's belief and denote the individuals $g(x_1), \ldots, g(x_n)$. Let R be formed by substituting $t_1, \ldots, t_n$ for free occurrences of $x_1, \ldots,$ x_n in Q. Then R is the sentence that (the speaker claims) John believes.

To refer to R, we must be able to refer to $t_1, \ldots, t_n$. For this purpose we choose new variables $y_1, \ldots, y_n$, distinct from $x_1, \ldots, x_n$. We will use the variable y_i to refer directly to the closed term t_i. Therefore we define a new assignment of values to variables. Let f be a function whose domain consists of the variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, with $f(x_i) = g(x_i)$ for i from 1 to n, and $f(y_i) = t_i$ for i from 1 to n.

If the free variables $y_1, \ldots, y_n$ have the values given by f, then the term subst('Q, ['$x_1, \ldots,$ 'x_n], [$y_1, \ldots, y_n$]) denotes the sentence R. The sentence "John believes that P" can express a singular proposition (S,f), where S is the following wff:

24 denotes$(y_1, x_1) \wedge \cdots \wedge$ denotes$(y_n, x_n) \wedge$ believe(john, subst('Q, ['$x_1, \ldots,$ 'x_n], $[y_1, \ldots, y_n]$)).

This proposition says that for i from 1 to n, t_i is a closed term of representation language that denotes the individual $g(x_i)$, and John believes the sentence R formed by substituting each t_i for the variable x_i in the wff Q. Since $x_1, \ldots, x_n$ includes all free variables of Q, and there are no free variables in $t_1, \ldots, t_n$, the result of this substitution is indeed a sentence, not a wff—just as the sentential theory requires. Note that the variables $x_1, \ldots, x_n$ have both quoted and unquoted occurrences in this wff. These variables were already present in the singular proposition (Q,g) expressed by the sentence P, where they referred directly to the individuals $g(x_1), \ldots, g(x_n)$. The wff Q appears under a quotation mark in wff (24), but the occurrences of $x_1, \ldots, x_n$ under the quotation mark do not refer directly. Instead they serve to indicate where the closed terms $t_1, \ldots, t_n$ occur in John's belief. For i from 1 to n, the variable x_i also occurs unquoted in (24), as the second argument of the predicate "denote." The unquoted occurrence of x_i refers directly to the individual $g(x_i) = f(x_i)$. We could avoid this double use of the variables $x_1, \ldots, x_n$ by complicating the semantic interpretation rule.

Notice that $t_1, \ldots, t_n$ can be any closed terms of the representation language. For each choice of $t_1, \ldots, t_n$ we get a possible interpretation for an utterance of the sentence "John thinks that P." This interpretation is "possible" in the sense that it is consistent with the syntax and semantics of English. The speaker of the utterance intends to refer to a particular sequence of terms $t_1, \ldots, t_n$, and the hearer must use knowledge of the speaker and situation to make some inferences about $t_1, \ldots, t_n$. It is possible that the hearer will infer enough about $t_1, \ldots, t_n$ to identify each of them uniquely, but we should not assume that the hearer must actually identify $t_1, \ldots, t_n$. It is possible that the hearer can learn something from the utterance, or produce a helpful response, given only a limited knowledge about the terms $t_1, \ldots, t_n$ that the speaker intends to refer to.

The object of an attitude verb may itself be an attitude report: for example, "I think you think you are smart." Let us check that our proposal allows a reasonable semantic interpretation for this example. Suppose the speaker is John and the hearer is Mary. John knows that in Mary's belief about her own smartness, the term that denotes Mary is her selfname. He does not know which constant is Mary's selfname, but he can refer to that constant as selfname(mary). Then John's belief about Mary's belief would be

25 bel(mary,subst('smart(x), ['x], [selfname(mary)]).

According to our semantics, an utterance of "I think you think you are smart" has many possible interpretations. I will show that there is one possible interpretation that is true iff John believes (25).

The clause "you are smart" expresses a singular proposition consisting of wff

26 smart(x)

and a function that assigns Mary as the value of the variable x. According to our semantics, the clause "you think you are smart" expresses a singular proposition consisting of the wff

27 denotes$(n, x) \wedge$ think(y,subst('smart(x), ['x], [n]))

and a function f that assigns values to the free variables of this wff. The variable x occurred in wff (26), and the function f assigns it the same value that it had in (26):

$f(x)$ is Mary. The variable y represents the pronoun "you," and the hearer is Mary, so $f(y)$ is also Mary. The variable n could refer to any closed term denoting Mary. Let n refer to Mary's selfname; then $f(n)$ is Mary's selfname. We have now chosen one of the possible interpretations for an utterance of "you think you are smart."

The interpretation of "I think you think you are smart" will consist of the wff

28 $\text{denotes}(m_1, x) \wedge \text{denotes}(m_2, y) \wedge \text{denotes}(m_3, n)$
$\wedge\ \text{think}(z, \text{subst}(\text{'}(\text{denotes}(n, x) \wedge \text{think}(y, \text{subst}(\text{'smart}(x), [\text{'}x], [n]))),$
$[\text{'}x, \text{'}y, \text{'}n],$
$[m_1, m_2, m_3]\ )\ ).$

and a function g that assigns values to the free variables of this wff. The variable z represents the pronoun "I," so $g(z)$ is John himself. The variables x, y, and n occurred in the wff (27), so the function g assigns them the same values that f assigned. Since x and y refer to Mary, $g(m_1)$ and $g(m_2)$ must be closed terms that denote Mary. Let $g(m_1) = g(m_2) =$ the constant mary. Since n refers to Mary's selfname, $g(m_3)$ must be a closed term that denotes Mary's selfname. Let $g(m_3)$ be the term selfname(mary). We have now chosen one of the possible interpretations for an utterance of "I think you think you are smart."

To find the belief described by (28), we take the first argument of "subst" and we substitute the constant "mary" for x and y and the term "selfname(mary)" for n. The result is

29 denotes(selfname(mary),mary) $\wedge$ think(mary,subst('smart(x), ['x], [selfname(mary)]))

If John believes (29), he can trivially infer (25). Conversely, suppose John believes (25). It is common knowledge that an individual's selfname denotes that individual, so John can easily infer the conjunction (29). So we have found a possible interpretation of the utterance that is true iff John believes (25).

5. Conclusions

We are given a belief report: "Mary believes that *P*." We have a parser and semantic interpreter that can identify the proposition expressed by *P*. We want to use this information to predict Mary's behavior—say, her answer to a given question. Suppose we know the algorithm that Mary uses for answering questions. Since the algorithm's input is a set of representations, we need to find Mary's internal representation for the proposition expressed by the clause *P*.

If the proposition expressed by a clause is a sentence of a representation language, our problem is already solved: Mary's representation of the proposition just *is* the proposition. Unfortunately, we have seen that this kind of theory cannot explain opaque indexicals. It seems that Mary's representation must be distinct from the proposition it represents, and if we are given the proposition it will take some reasoning to find the representation. If we are going to distinguish between a proposition and its internal representation, we would like a clear and simple picture of how they are related, and how we are to find the representation when the proposition is known. We have proposed that the proposition expressed by *P* is a pair, containing a wff and an assignment of values to the free variables of the wff. To find Mary's representation we must find the terms she uses to represent the values of the free variables, and substitute these terms for the corresponding variables in the wff. In Crimmins and Perry's work, a proposition and its representation are two quite different objects. A proposition consists of objects and relations, while a representation consists of "ideas" and "notions." If we need both internal representations and singular propositions, it is simpler to assume that they are closely similar objects—instead of radically different objects, as in Crimmins and Perry.

The examples that motivate this work are unfortunately remote from the texts that computational linguists currently work on. In studying opacity we concentrate on cases where an agent uses two different representations without realizing that they refer to the same entity. Such cases are popular in fiction, but rare in applications like machine translation or question-answering from a database. So it will be some time before these ideas are helpful in the practice of computational linguistics. However, indexicals occur constantly in human speech, and opacity is a crucial issue in any theory of propositional attitudes. So if we are serious about a sentential theory of attitudes, it is important to be certain that such a theory can explain opaque indexicals.

Acknowledgment

This work was supported by the National Science Foundation under grant number IRI-9006832.

References

Barwise, Jon, and Etchemendy, John (1987). *The Liar*. Oxford University Press.

Barwise, Jon, and Perry, John (1983). *Situations and Attitudes*. MIT Press.

Castañeda, Hector-Neri (1968). "On the logic of attributions of self-knowledge to others." *Journal of Philosophy*, 65(15), 439–456.

Crimmins, Mark, and Perry, John (1989). "The prince and the phone booth: Reporting puzzling beliefs." *Journal of Philosophy*, 86(12), 685–711.

des Rivières, J., and Levesque, Hector (1986). "The consistency of syntactical treatments of knowledge." In *Theoretical Aspects of Reasoning about Knowledge*, edited by Joseph Y. Halpern, 115–130. Morgan Kaufmann.

Haas, Andrew (1986). "A syntactic theory of belief and action." *Artificial Intelligence*, 28(3), 245–292.

Haas, Andrew (1991). "Sentential semantics for propositional attitudes." *Computational Linguistics*, 16(4), 213–233.

Kaplan, David (1975). "Quantifying in." In *The Logic of Grammar*, edited by Donald Davidson and Gilbert Harman, 112–144. Dickinson.

Kaplan, David (1989). "Demonstratives." In *Themes from Kaplan*, edited by Joseph Almog, John Perry, and Howard Wettstein, 481–563. Oxford University Press.

Konolige, Kurt (1986). *A Deduction Model of Belief*. Pitman.

Kripke, Saul (1975). "Outline of a theory of truth." *Journal of Philosophy*, 72(9), 690–715.

Montague, Richard (1974). "Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability." In *Selected Papers of Richard Montague*, edited by Richmond H. Thomason. Yale University Press.

Moore, Robert C., and Hendrix, Gary G. (1982). "Computational models of belief and semantics of belief sentences." In *Processes, Beliefs, and Questions*, edited by S. Peters and E. Saarinen, 107–127. D. Reidel.

Perlis, Donald (1988). "Languages with self-reference II: Knowledge, belief, and modality." *Artificial Intelligence*, 34(2), 179–212.

Reichenbach, Hans (1947). *Elements of Symbolic Logic*. The Free Press.