COGNITIVE SCIENCE **10**, 371–422 (1986)

Logical Foundations for Belief Representation

WILLIAM J. RAPAPORT
State University of New York, Buffalo

This essay presents a philosophical and computational theory of the representation of *de re*, *de dicto*, nested, and quasi-indexical belief reports expressed in natural language. The propositional Semantic Network Processing System (SNePS) is used for representing and reasoning about these reports. In particular, quasi-indicators (indexical expressions occurring in intentional contexts and representing uses of indicators by another speaker) pose problems for natural-language representation and reasoning systems, because—unlike pure indicators—they cannot be replaced by coreferential NPs without changing the meaning of the embedding sentence. Therefore, the referent of the quasi-indicator must be represented in such a way that no invalid coreferential claims are entailed. The importance of quasi-indicators is discussed, and it is shown that all four of the above categories of belief reports can be handled by a single representational technique using belief spaces containing intensional entities. Inference rules and belief-revision techniques for the system are also examined.

This essay presents a computational analysis of a referential mechanism—quasi-indexicality—first examined in philosophy some 20 years ago, but not hitherto employed in artificial intelligence (AI) studies of belief systems. In turn, a philosophical claim about the relations of *de re, de dicto,* and *de se* beliefs is made as a by-product of the computational analysis. I thus hope to illustrate the importance of philosophy for research in AI and the correlative importance of a knowledge of AI for philosophical research, in the spirit of Dennett's (1978) recommendations:

This research was initially carried out in collaboration with Stuart C. Shapiro and has been supported in part by Research Development Fund Award No. 150-9216-F from the Research Foundation of State University of New York, and by the National Science Foundation under Grant No. IST-8504713. Earlier versions of this paper have been presented at SUNY Buffalo, SUNY Fredonia, Tulane University, COLING-84, and at St. Bonaventure University. I wish to express my gratitude to the SUNY Fredonia Department of Philosophy for a leave of absence that enabled me to begin the work culminating in this research. I am grateful to Hector-Neri Castañeda, Stuart C. Shapiro, and SNePS Research Group (especially João Martins, Ernesto Morgado, Terry Nutter, and Janyce M. Wiebe), and my colleagues in the University at Buffalo Graduate Group in Cognitive Science/Deictic Center Research Project (Gail A. Bruder, Judith Duchan, Erwin M. Segal, Stuart C. Shapiro, and David A. Zubin) for comments and criticism. Some of this research was first reported in Rapaport and Shapiro (1984) and Rapaport (1984b).

Correspondence and requests for reprints should be sent to William J. Rapaport, Department of Computer Science and Graduate Group in Cognitive Science, SUNY, Buffalo, NY 14260.

> Philosophers, I have said, should study AI. Should AI workers study philosophy? Yes, unless they are content to reinvent the wheel every few days. When AI reinvents a wheel, it is typically square, or at best hexagonal, and can only make a few hundred revolutions before it stops. Philosopher's wheels, on the other hand, are perfect circles, require *in principle* no lubrication, and can go in at least two directions at once. Clearly a meeting of minds is in order. (p. 126)

1. OVERVIEW

1.1. Belief Representation versus Knowledge Representation

The branch of AI and computer science known as 'knowledge representation' is concerned with how to represent "knowledge of the application environment. . . and knowledge of the intended audience" in a computer system (cf. McCalla & Cercone, 1983, p. 12). To a philosopher, the name of this field is fundamentally misleading. A more neutral term would be something like 'information representation' or 'data representation.' When knowledge representation is taken to be more generally computer-science-oriented than AI-oriented, either of these terms would be more appropriate.

But when it is taken as a subject concerned with the representation of information in AI systems, its epistemic connotation comes to the fore. In this context, however, more than mere knowledge is being represented. For what is represented are such things as objects, properties, situations, and propositions. Although these *can* be the objects of knowledge (one can know an object, or know what it is; one can know what properties it has; one can know that a certain proposition is true, etc.), they are, more generally, the objects of *belief* and *acquaintance*—in general, the objects of *thought*.

The distinction between knowledge, in particular, and beliefs or thoughts, in general, is an important one, for one can think about things that do not exist and one can believe propositions that are, in fact, false (cf. Meinong, 1904/1971; Rapaport, 1978). But one cannot *know* a false proposition. Yet, if an AI system is to simulate (or perhaps *be*) a mind or merely interact with humans, it must be provided with ways of representing nonexistents and falsehoods. Because belief is a part of knowledge and has a wider scope than knowledge, the term *belief representation* is a more appropriate one in an AI context.

1.2. A Belief-Representation System

The ultimate goal of the research described here is the construction of an AI system that can reason about the beliefs and other cognitive states of intelligent agents, including humans (e.g., the system's users), other AI systems (e.g., interacting ones), and itself (cf. Nilsson, 1983, p. 9). The cognitive states include goals, intentions, desires, hopes, and knowledge, in addition to beliefs. Here, I shall be concerned only with beliefs, partly because

belief plays a central logical and psychological role in the network of cognitive states, partly because it is in many ways simpler to analyze than other cognitive states, and partly to be able to build on the large philosophical and computational literature about belief.

The sort of AI system that is of concern here is a representation and reasoning system whose "data base" contains information about the world and about various cognitive agents. In order for the system to learn more about these agents (and the world)—to expand its "beliefs"—it should contain information about *their* beliefs and be able to reason about them. Such a data base constitutes the beliefs of the system about these agents and about their beliefs.

Because each of the agents is in fact such a system itself, each has beliefs about the beliefs of the others. Thus, the system must be able to represent (i.e., have beliefs about) beliefs about beliefs and to reason about these. Such beliefs are often referred to as *nested* beliefs.

A belief-representation system must also be sensitive to the *intensionality* of belief and to the associated phenomenon of *referential opacity*. I shall have more to say about this in Section 4. For now, it suffices to note two important points.

First, the intensionality of belief puts constraints on the system's inference mechanism. For instance, given the system's beliefs that an agent, A, believes some proposition p and that p is logically equivalent to another proposition q, the system should not infer that A believes q, in the absence of further information.

Second, an agent can have inconsistent beliefs about an object. For instance, A might believe both that the Evening Star is a planet and that the Morning Star is not a planet, even though the Morning Star *is* the Evening Star. This can happen as long as A does not believe that the Morning Star is the Evening Star. In this case, A's "data base" contains *two* items, one for the Morning Star, one for the Evening Star. Such items are *intensional objects*, and our system must be able to deal with them.

The focus of this essay is a discussion of another requirement for such a system—one that has not been discussed in previous computational literature: The system must be sensitive to the *indexicality* of certain beliefs, in particular, to the phenomenon of *quasi-indexicality*. This, too, will be dealt with in much greater detail later (in Section 3). Briefly, it is a feature that is at the core of *self-referential* beliefs—that is, beliefs about oneself—and their expression by others. Thus, the belief that A would express by "I am rich" must be reported by someone else thus: "A believes that *he* is rich"; it clearly should not be reported as "A believes that I am rich."[1]

Finally, the system ought to be able to expand and refine its beliefs by interacting with users in ordinary conversational situations. This is not strictly

[1] Nor should it be reported as "A believes 'I am rich'." For arguments to this effect, see Church (1950), Feldman (1977), and Cresswell (1980).

a requirement. Users could be required to learn a rigid, canonical language for unambiguously expressing beliefs and to use this language with the system. Indeed, the system described here has this requirement. But if the system is to be considered as a cognitive agent, and especially if it is to be used as a tool in understanding *our* belief-representation mechanisms, it ought to interpret ordinary statements about belief, expressed in (grammatical) natural language, the way humans do. Thus, we would want the system to make reasonable or plausible interpretations of users' belief reports—based on such things as subject matter and prior beliefs (including beliefs about the user and the user's beliefs)—and to modify its initial representation as more information is received. Techniques for accomplishing this are considered in the final section of this essay.

1.3. Methodology

I shall begin by examining ways of representing two distinct kinds of belief reports (*de re* and *de dicto* reports) and the special report (a species of *de se* reports) involving quasi-indexical reference. These reports will be given in English using canonical representations and will be translated into a semantic-network representation. The translations have been implemented using the Semantic Network Processing System (SNePS) (Shapiro, 1979) and an Augmented Transition Network (ATN) parser-generator with a deductive question-answering capability (Shapiro, 1982). (The representation was first presented in Rapaport and Shapiro [1984]; details of the implementation were presented in Rapaport [1984b].)

The result is a computational study of the *logic of belief.* It is important to note that here I am concerned only with the *logic* of belief reports expressed in a canonical language; thus, I shall *not* be concerned with the *pragmatic* problems[2] of determining from context and prior belief *which* canonical representation was intended by a speaker. It is also important to note that here I am concerned only with the logic of *belief;* thus, I shall not deal with the issue of how different objects of belief—different topics of conversation—might affect the interpretation of a natural-language belief report. I view these issues roughly as issues of *performance,* as opposed to the issues of *competence* that, I feel, must be dealt with first. Once we see how to represent beliefs in clear cases, we can then turn to the more complex linguistic, psychological, and pragmatic issues of the interpretation of ordinary language.

2. THE IMPORTANCE OF BELIEFS

In this section, several arguments will be presented for the importance of representing and reasoning about beliefs in general, about nested beliefs in

[2] In one sense of that word. In another—the sense in which pragmatics is the study of indexicals—this study is *precisely* concerned with pragmatics.

particular, and (briefly) about beliefs involving quasi-indicators. I begin by considering three important roles that beliefs play, both for humans and in AI systems: *evidentiary, exploratory,* and *behavior-producing roles.*[3]

2.1. The Evidentiary Role of Beliefs

One's own beliefs, as well as the beliefs of others, can be used as evidence or premises for coming to believe propositions. As an example, to decide whether I should believe *p,* I might reason as follows:[4]

John believes *q*.
I believe what John believes.
I believe that *p* is logically equivalent to *q*.
Therefore, I should believe *p*.

Thus, if an AI system is going to be able to increase and refine its data base —its "beliefs"—it ought to be able to represent and reason about its own beliefs as well as those of its users.

2.2. The Explanatory Role of Beliefs

Actions can be either intended or unintended. That is, an action can be the result of a (conscious) decision, or it might be merely accidental (a "behavior"). In the former case, it would be the end result of a chain of reasoning (a "plan") that would include beliefs. An AI system (perhaps an intelligent robot) that would be capable of performing actions, or even a system that would be capable of recommending actions to its users, would thus need to be able to deal with beliefs. Such a system would need to deal with nested beliefs if its plans would have to take into account other agents and their plans (which might either aid or hinder the system's plans).

When a belief is a cause of an agent's actions, we are interested not only in *what* the agent believes, but also in *how* the agent believes it. That is, we are interested not only in a third-person characterization of the agent's beliefs, but also in the agent's *own* characterization of those beliefs. The distinction between these two ways of reporting beliefs is captured by means of the distinction between *de re* and *de dicto* belief representations. Thus, an AI system that is capable of explaining or recommending behavior must be able to distinguish between these two kinds of belief reports by having two distinct means of representing them. (These two kinds of belief reports are described in more detail in Section 4.3.)

2.3. The Behavior-producing Role of Beliefs

Several AI systems (e.g., HAM-ANS [Wahlster, 1984; Marburger, Morik, &

[3] The first two of these were stressed by John Perry in his lectures on Situation Semantics during the COLING-84 Summer School, Stanford University, 1984. Some of the points and examples are adapted from these lectures.

[4] For an interesting philosophical discussion of this sort of belief justification, see Hardwig (1985).

Nebeli, 1984], GRUNDY [Rich, 1979], UC [Wilensky, Arens, & Chin, 1984]) employ the technique of *user modeling* in order to produce appropriate natural-language output. An interactive dialogue system can build a profile of each user in order to tailor its output to the user's needs. This profile might include such information as: (a) the user's degree of familiarity with the topic (as in UC, where the naive user must be distinguished from the expert), (b) the user's current "state of knowledge"—more accurately, the user's current set of beliefs, whether or not they correctly reflect the world, (c) the user's interests, and (d) the mutual (or shared) beliefs of the system and user (in order to be able to deal with such questions as whether they are talking about the same objects or have the same beliefs) (cf. Wahlster, 1984).

Not only would such a representation of the user's beliefs (and other cognitive attitudes) be used in determining the *level* of the dialogue (e.g., whether the system's output is to be used by a novice or an expert), but also in determining the *quality* and *quantity* of the information output (e.g., whether the system should provide only literal answers, whether it should provide more discursive ones that give slightly more—or less—information in order not to be misleading [cf. Joshi et al., 1984]).

2.4. Belief Spaces

The discussions in the previous sections lead naturally to the notions of *belief spaces* (Martins, 1983) or *views* (Konolige, 1985). The central idea behind these notions is the set of beliefs of an agent. In this essay, I consider the set of propositions currently believed by an agent, together with the set of items about which the agent has beliefs, *relativized to a "reporter" of these beliefs,* to be the *belief space of the agent in the context of the "reporter."*

The items about which an agent has beliefs need not exist; they might be distinct but co-extensive—as in the Morning Star/Evening Star case—and they might be the same as, or different (or differently represented) from, the items in another agent's belief space. Moreover, with the possible exception of the system's own beliefs, the belief space of an agent *as represented by the system* will contain *not* the agent's (own) representations of the objects of his beliefs, but the *system's* representations of the agent's representations of them. And, in the case of nested beliefs, the objects of an agent's beliefs would be represented not as the agent represents them, but as another agent would represent (report) them. Thus, some care must be taken in the choice of a representational scheme for an AI system that can reason about beliefs.

2.5. The Importance of Nested Beliefs

Discussions of nested beliefs such as

(1) John believes that Mary believes that Lucy is rich

occasionally produce the observation that the speaker must be joking, that no one really talks that way. Although there are, no doubt, performance

limitations on the allowable depth of such nesting, our linguistic competence clearly allows for the grammatical (if not acceptable) formulation of such sentences. Any sentence can be prefixed by an operator of the form '*A* believes that' (where *A* names a cognitive agent). Indeed, Kant held that *all* statements were within the scope of an implicit 'I think that' operator (Kant, 1787/1929, p. B131), so even as simple a statement as

John believes that $1 + 1 = 2$.

is really of the form

I (the speaker) think that John believes that $1 + 1 = 2$.

But the performance limitation must be taken seriously. The feeling that a sentence such as (1) must be a joke raises the question of how deep the nesting actually can be in ordinary cases (cf. Dennett, 1983, p. 345).

There are some clear cases where up to three occurrences of 'believes that' are natural and of importance in explaining behavior. For instance, as John Perry has pointed out, each of the participants in the Situation Semantics course at COLING-84 attended because of their beliefs about Perry's beliefs, as well as because of their beliefs about the other student's beliefs about Perry's beliefs (e.g., that the other students believed that the theory was worthwhile, hence I ought to believe it, too).

Another natural case is the following: Suppose that I tell Mary (truthfully) that John thinks that she doesn't like him. In that case, I believe that John believes that Mary believes that he is dislikable. (Actually, I tell Mary that I believe that John believes that Mary believes that he is dislikable; here, there are *four* levels of nesting of propositional attitudes.)

Finally, nested propositional attitudes also occur when one considers the structure of an information-seeking dialogue. Here, a user might come to an AI system in the hopes that it will provide information. If the user is not clear about the exact nature of the information needed, the attitude of the system must be to *wonder* about what the user *wants to know* (cf. Carberry, 1984).

I shall have more to say about nested beliefs, and especially their interaction with quasi-indicators, in Section 3.3.1.

2.6. Pronominal Reference

One of the more difficult issues in natural-language understanding is how to determine the reference of pronouns. In *anaphoric* reference, a pronoun's antecedent occurs within the text (e.g., 'John is tall, and *he* is clever'). In *indexical* reference, a pronoun refers to an entity outside of the text (e.g., '*She* [i.e., that woman over there] is a philosopher'). There is an important case of pronominal reference in which the pronoun occurs within a belief (or other intentional) context (i.e., within the scope of a 'believes that' or other intentional-attitude operator). In this case, the pronoun refers to an entity

outside of its own "level" or "nesting" of the text, yet its antecedent occurs within the text. An example is the occurrence of 'he' in

(2) John believes that he is rich.

This is a *quasi-indexical* use of 'he' (it might also be—but is not—called 'quasi-anaphoric'), and it behaves differently from both pure anaphoric and pure indexical uses (as we shall see in Section 3.1). Yet no other AI system that deals with beliefs is fully sensitive to its unique qualities. One of the main purposes of this essay is to show how an AI system for reasoning about beliefs can, and must be able to, handle quasi-indexical reference. I now turn to the promised discussion of that phenomenon.

3. QUASI-INDICATORS

3.1. The Nature of Quasi-Indexical Reference

Beginning in the 1960s, Hector-Neri Castañeda wrote a series of papers in which he introduced and elaborated the theory of quasi-indexical reference (Castañeda, 1966, 1967a, 1967b, 1968a, 1968b, 1970, 1975c, 1980, 1983; Adams & Castañeda, 1983). The theory has been subjected to a great deal of philosophical scrutiny, but it has generally emerged unscathed, and its importance has never been questioned (cf. e.g., Hintikka, 1967; Perry, 1979, 1983, 1985; Lewis, 1979; Boër & Lycan, 1980; Stalnaker, 1981; Brand, 1984).

Following Castañeda (1967b, p. 85), an *indicator* is a personal or demonstrative pronoun or adverb used to make a strictly demonstrative reference, and a *quasi-indicator* is an expression within an intentional context (typically, one within the scope of a verb or propositional attitude, such as 'believes that') that represents a use of an indicator by another person.

Suppose that person A says to person B at time t and place p, "I am going to kill you here now." Suppose further that person C overhears this, calls the police, and says, "A said to B at p at t that he* was going to kill him* there* then*." The starred words are quasi-indicators representing uses by A of the indicators, 'I', 'you', 'here', and 'now', as reported by C.

There are two properties (among many others) of quasi-indicators that must be taken into account when making representation decisions for our AI system:

- Quasi-indicators occur only *within* intentional contexts.
- Quasi-indicators cannot be replaced, preserving truth value, by any co-referential expressions.

The general question they are intended to help answer is: "How can we attribute indexical references to others?" (Castañeda, 1980, p. 794).

The specific cases that I am concerned with are exemplified in the following scenario. Suppose that John has just been secretly appointed editor of

Cognitive Science, but that John does not yet know this. Further, suppose that, because of a well-publicized salary accompanying the office of *Cognitive Science*'s editor, which is traditionally immediately deposited in the new editor's account,

(3) John believes that the editor of *Cognitive Science* is rich.

And suppose finally that, because of severe losses in the stock market,

(4) John believes that he himself is not rich.

Suppose that the system had information about each of the following: John's appointment as editor, John's (lack of) knowledge of this appointment, and John's belief about the wealth of the editor. We would *not* want the system to infer

(2) John believes that he* is rich.

because this is inconsistent with (4), which is consistent with the rest of the system's information. The 'he himself' in (4) is a quasi-indicator, for (4) is the sentence that *we* would use to express the belief that *John* would express as 'I am not rich'. Someone pointing to John, saying,

(5) He [i.e., that man there] believes that he* is not rich.

could just as well have said (4). The first 'He' in (5) is not a quasi-indicator: It occurs outside the believes-that context, and it can be replaced by 'John' or by 'the editor of *Cognitive Science*', preserving truth value. But the 'he*' in (5) and the 'he himself' in (4) could not be thus repeated by 'the editor of *Cognitive Science*'—given our scenario—even though John *is* the editor of *Cognitive Science.* And if poor John also suffered from amnesia, it could not be replaced by 'John', either.

3.2. The Importance of Quasi-Indexical Reference

Clearly, a system capable of reasoning about an agent's beliefs must be able to handle quasi-indicators if it is not to draw faulty conclusions. Moreover, theories that do not take quasi-indexical reference into account do so at the expense of being unable to represent an important category of beliefs, namely, beliefs about oneself.[5] And a number of philosophers, from John Perry (1979) to, most recently, Myles Brand (1984), have emphasized the importance of such beliefs for explaining and producing actions.

3.2.1. Quasi-Indicators and Action. Suppose that Mary is the tallest woman in Muir Woods (at some time t), suppose that a large tree is about to fall on her, and suppose that she comes to believe this (say, because someone

[5] Such beliefs also play a role in purely philosophical speculation, such as in discussions of the Cartesian *cogito* (cf. Rapaport 1976a, pp. 63, 67 n.1).

tells her). The propositional "content" of her belief—that is, the internal make-up of the object of her belief—will determine whether or not she takes evasive action. (More precisely, it will be of use in explaining her subsequent behavior, whether evasive or not.) To see why, consider the following two reports of Mary's belief state:

(6) Mary believes that a tree is about to fall on the tallest woman in Muir Woods.

(7) Mary believes that a tree is about to fall on her*.

Mary is unlikely to take evasive action if she believes what (6) reports her as believing, *unless she also believes that she* is the tallest woman in Muir Woods.* But those two beliefs taken together would (normally) produce in Mary the belief state reported in (7). Both the belief in case (7) and the additional belief required for action in case (6) contain quasi-indicators in their third-person reports ('her*' in the former, 'she*' in the latter).

3.2.2. Other Belief-Representation Systems. In this section, we consider other belief-representation systems and their (lack of) treatment of quasi-indexical reference. (For a more thorough survey, see Rapaport [in press].)

3.2.2.1. Moore. One of the first AI researchers to recognize the importance of an AI system capable of reasoning about knowledge was Robert C. Moore.

> AI systems need to understand what knowledge [they and the systems or people they interact] with have, what knowledge is needed to achieve particular goals, and how that knowledge can be obtained. (Moore, 1977, p. 223)

Moore also recognized

> how intimately the concept of knowledge is tied up with action. . . . [T]he real importance of such information is usually that it tells us something about what . . . [the agent] can do or is likely to do. (Moore, 1977, p. 223)

Note, however, his talk of *knowledge* rather than of *belief*. "Surely," he says, "one of the most important aspects of a model of another person is a model of what he knows" (Moore, 1977, p. 223). But a model of what he *believes* is far more important: For it is people's *beliefs*—including their *knowledge*, as well as their *mistaken* beliefs—that enable them to have goals and perform actions.

In a later work, Moore points out that it is not "clear that knowledge can be *defined* in terms of belief" (Moore, 1980, p. 33; italics added). Although this is true, it *is* clear, however, that knowledge implies true belief, so belief is still the more fundamental concept. This is, in fact, assumed by Moore when he notes that "knowledge tends to be cumulative, belief does not" (Moore, 1980, p. 34)—the point is that what's known is believed and is true, hence there are no problems with belief *revision*. This is, no doubt, a problem; we consider it in Section 6.

Moore also recognized the importance of-being able to deal with referentially opaque contexts, even though his method for so dealing (use of quotation) leaves something to be desired (cf. Section 1.2, fn. 1). He recommended the use of Hintikka's (1962) logic of knowledge; although this is far superior to, say, (re)inventing his own such logic, buying into someone else's theory carries certain risks, notably the inability—well-known at the time—of Hintikka's system to deal with quasi-indicators (cf. Castañeda, 1966, p. 130, fn. 1; Castañeda, 1967a, p. 11ff; Hintikka, 1967). Finally, Moore (1980, p. 87) does show an awareness of the quasi-indicator problem, but, rather than *solving* the problem by showing how to represent quasi-indicators, he *avoids* the problem by using a rigid designator to denote the cognitive agent. This, however, works (at best) only if one is willing to accept the theory of rigid designators and to embed one's theory in the context of possible worlds. The theory that I am putting forth, however, is purely intensional in the sense of not using—or needing—possible worlds or, hence, rigid designators (see Section 4.1.3).

3.2.2. McCarthy. John McCarthy (1979) emphasized the importance of using what he called *individual concepts*[6] (a kind of intensional entity; see Section 4.1.3) in computational analyses of knowledge. However, his intensional stance was not very thoroughgoing, because he mixed intensional and extensional entities together. Like Moore, he also emphasized the importance of knowledge for action: "A computer program that wants to telephone someone must reason about who knows the number. More generally, it must reason about what actions will obtain needed knowledge (McCarthy, 1979, p. 145). Presumably, it must also reason about what *knowledge* will lead to desired *actions.* But, as with Moore, he concentrated on *knowledge* rather than belief. Nor, as is by now to be expected, does he show how to deal with quasi-indicators.

3.2.2.3. Cohen and Perrault. Philip R. Cohen and C. Raymond Perrault's work on speech acts have axioms for belief that are for an idealized believer (Cohen & Perrault, 1979, p. 480, fn. 5). They recognize the importance of belief for action, the importance of nested beliefs in such communicative acts as slamming a door (the slammer intends that the observer believes that the slammer intends to insult him); and the consequent importance of representing the *agent's* model of *another* agent's beliefs (Cohen & Perrault, 1979, p. 478, 480).

But they miss the need for quasi-indicators. Their axiom B.2 (Cohen & Perrault, 1979, p. 480, fn. 5) is:

aBELIEVE(P)→aBELIEVE(aBELIEVE(P))

[6] This term, like many others in logical and philosophical studies of intensionality, means different things to different authors (cf. Section 4.1.3.1.). Because most computational researchers have not tried to make their uses of such terms precise, I shall not examine them in detail.

but the third occurrence of 'a' should be a quasi-indicator, not 'a' (cf. my discussion of Maida and Shapiro's [1982] Rule 2 in Section 5.1). Similarly, their definitions of such 'operators' as:

MOVE(AGT, SOURCE, DESTINATION)	
CANDO.PR:	LOC(AGT, SOURCE)
WANT.PR:	AGT BELIEVE AGT WANT move-instance
EFFECT:	LOC(AGT, DESTINATION)

have a WANT.PRecondition that requires that the AGenT know his own name: The second occurrence of 'AGT' should be a quasi-indicator.[7] What is of significance here is the importance of quasi-indicators in *nested* belief contexts.

3.2.2.4. Schank. Although Roger Schank's conceptual dependency theory was not intended as a belief-representation system, it also fails to take note of the importance of quasi-indicators. The primitive actions MTRANS and MBUILD are the ones that need quasi-indicators as slot fillers. For instance, one CD analysis of

(8) John promised to give Mary a book.

is (cf. Schank & Riesbeck, 1981, pp. 19–20):

(8CD)	actor	:	John		
	action	:	MTRANS		
	object	:	actor	:	John
			action	:	ATRANS
			object	:	book
			to	:	Mary
			from	:	John
	from	:	John		

But this is not fine-grained enough. A similar analysis of

John promised that Lucy would give Mary a book.

would be something like:

actor	:	John		
action	:	MTRANS		
object	:	actor	:	Lucy
		action	:	ATRANS
		object	:	book
		to	:	Mary
		from	:	Lucy
from	:	John		

[7] As Stuart C. Shapiro has pointed out to me, this assumes that 'a' and 'AGT' are names.

But without a quasi-indicator as the filler of the "object : actor" slot, (8CD) is the analysis, not of (8), but of

(9) John promised that John would give Mary a book.

And, as our study of quasi-indicators has shown, (8) and (9) are not equivalent. (A similar argument against Schank is given in Brand [1984, pp. 209–212].)

3.2.2.5. Clark and Marshall. The work of Herbert H. Clark and Catherine R. Marshall (1981) also points up—by its absence—the importance of quasi-indicators in nested belief contexts. They emphasize the importance of "shared knowledge" (which, incidentally, requires the use of some sort of "coreferentiality" mechanism across belief spaces) and, in particular, of "mutual belief": a sort of infinite nesting of beliefs (Clark & Marshall, 1981, p. 12).

But by not taking quasi-indicators into account, their analysis of the requirements for the use of definite referring expressions is incomplete. For instance, their third clause is:

Ann knows that Bob knows that Ann knows that t is R.

But they do not point out that this also requires the condition that

Ann knows that Bob knows that she herself is Ann.

The 'she herself' is a quasi-indicator. A simpler alternative, one not needing this extra belief but still needing a quasi-indicator, is:

Ann knows that Bob knows that she herself knows that t is R.

3.2.2.6. Wilks and Bien. Yorick Wilks and Janusz Bien argue "that there can be a very general algorithm for the construction of beliefs about beliefs about beliefs" (Wilks & Bien, 1983, p. 96), but feel

> that a belief manipulating system, which is to be psychologically and computationally plausible, must have built into it some limitations on processing, so as to accord with the fact that deep nestings of beliefs (while well formed in some "competence sense") are in practice incomprehensible. (Wilks & Bien, 1983, p. 97)

They explicitly distance themselves from "logic-based approaches" (Wilks & Bien, 1983, p. 97). Nevertheless, they are concerned with a number of representional issues of importance to us. For instance, they make it clear that one must distinguish between an agent's beliefs as the *agent* would represent them, and the agent's beliefs as the *system* represents them; they are especially concerned about dealing with this complication in the case of nested beliefs.

They are also concerned with what they call *self-embedding* (Wilks & Bien, 1983, pp. 114–117): dealing with the system's beliefs about itself or the system's beliefs about a user's beliefs about himself. Yet there is no indication of how their system would handle quasi-indexical reference and this seems especially damaging to their theory: They give an example of the construction of the system's view of a user's view of a person by "pushing" the system's view of the person "down into" the system's view of the user (Wilks & Bien, 1983, pp. 104–107). Suppose that, as in our story, John is the editor of *Cognitive Science* (but doesn't know it), John doesn't believe that he* is rich, and John believes that the editor of *Cognitive Science* is rich. In Wilks and Bien's notation, the system's view of John is as shown in example (I):

(I)

John
* is not rich.
The editor of *Cognitive Science* is rich.
---
* = the editor of *Cognitive Science*
* is rich.
The editor of *Cognitive Science* is rich.

system

where '*' refers to John, the beliefs above the line are the sytem's beliefs about John's beliefs, and the beliefs below the line are the system's beliefs about John.

Now, what happens when this is pushed down into *itself*? If I have followed their example correctly, there is an intermediate stage (see example (II)):

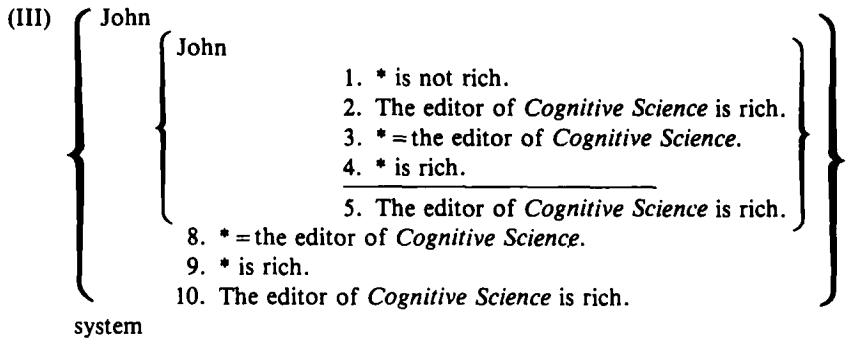
(II)

John
John
1. * is not rich.
2. The editor of *Cognitive Science* is rich.
---
3. * = the editor of *Cognitive Science.*
4. * is rich.
5. The editor of *Cognitive Science* is rich.

6. * is not rich.
7. The editor of *Cognitive Science* is rich.
---
8. * = the editor of *Cognitive Science.*
9. * is rich.
10. The editor of *Cognitive Science* is rich.

system

Then: Beliefs 3 and 4 "mount to the upper half" (Wilks & Bien, 1983, p. 106); beliefs 6 and (maybe) 7 "enter the inner" environment as copies of beliefs 1 and (maybe) 2; and beliefs 8, 9, and 10 similarly migrate as copies of beliefs 3, 4, and 5. The result is as shown in example (III):



But this is both incorrect (because John does not believe either 3 or 4) and contradictory (because 1 and 4 are contradictory). This is because sometimes '*' refers to the *system's* John and at other times it is the system's (quasi-indexical) mechanism for John's *self*-reference. Wilks and Bien get into trouble (if my analysis of this case is correct)[8] by ignoring the distinctions between these two uses of '*'.

3.2.2.7. Konolige. Kurt Konolige has been concerned, *inter alia,* with reasoning about the beliefs of others. He provides "a. . . formal model of belief . . . for representing situations in which belief derivation is logically incomplete" (Konolige, 1985, p. 360). In this model, each agent has his own set of inference rules; the system can reason about an agent's beliefs by using the agent's rules. Thus, it can distinguish between the beliefs an agent would have if the agent were "ideal" (i.e., if the agent believed all the logical consequences of his beliefs) from the beliefs that the agent actually has. (This technique, incidentally, has also been implemented in Martins [1983].)

Even though Konolige's theory is supposed to deal with realistic limitations on belief systems, he apparently identifies the system's beliefs with facts about the real world (Konolige, 1985, p. 385). But the system ought to be treated the same as any other agent: As I noted in Section 2.5, the system's beliefs are all implicitly prefixed by an 'I believe that. . .' and the objects of its beliefs are not items in the real world.

And, as with Wilks and Bien, it is not clear how Konolige's system would deal with quasi-indicators—if it deals with them at all. For instance, he says:

---

[8] It might not be correct. But either of two other plausible interpretations have similar problems: First, perhaps 4 moves up, *overwriting* 1; then 6 might move up to the bottom of the inner environment, move up to the top, and re-overwrite the moved-up 4. This would result in an inner environment with only beliefs 1, 2, and 3 (as in the last diagram); but this by itself is a contradictory set of beliefs if classical logic is used.

Alternatively, perhaps 6 moves up to the bottom of the inner environment *before* 4 moves up, overwrites 4, and then moves up to the top of the inner environment. This gives the same result.

> we might argue that, if an agent *S* believes a proposition *P*, then he believes that he[*] believes it. All he has to do to establish this is query his[*] belief subsystem with the question. "Do I believe *P*?" If the answer comes back "yes," then he should be able to infer that he[*] does indeed believe *P*, *i.e.* [*S*][*S*]*P* is true if [*S*]*P* is. (Konolige, 1985, p. 385; '[*S*]*P*' is Konolige's notation for '*S* believes *P*'. I have indicated quasi-indicators by '[*]')

But [*S*]*P* should not imply [*S*][*S*]*P*, since *S* might believe *P* and believe that *he** believes *P*, yet *not* believe that *S* believes *P*, because *S* might not believe that he* is *S*. It is also important that, when *S* queries a belief subsystem, he query *his own* subsystem, and not merely the subsystem of someone named '*S*'.[9]

3.2.2.8. Creary. A bit more philosophical subtlety is evident in the work of Lewis G. Creary (1979; Creary & Pollard, 1985). Creary (1979) points out that AI systems ought to reason about knowledge and belief for the simple reason that humans do. He cites four main problems that a belief system ought to be able to deal with: the philosophical problems of referential opacity, quantifying in (i.e., embedding belief reports in the scope of existential quantifiers), and nested beliefs; and the computational problem of making realistic inferences. His treatment seems to be a thoroughly intensional one: Knowledge and belief are taken to be relations between agents and *propositions,* and propositions appear to be constituted by relations holding among intensional entities, thus allowing him to quantify into belief contexts (cf. Creary, 1979, p. 177). He also *seems* to recognize that an item referred to within a nested belief context is not necessarily the same as an item referred to outside the context (Creary, 1979, p. 178), but his notation and explanations are far from clear. Similarly, he points to ambiguities in the expression of beliefs, but it is hard to tell if he has the *de re/de dicto* or some other distinction in mind.

One point that *is* clear, from his very first example, is the lack of sensitivity of his system to the problem of quasi-indicators: His analysis of 'Mike wants to meet Jim's wife as such' is:

> wants(mike, Meet{Mike, Wife Jim})

(Creary, 1979, p. 177). The capitalized 'Mike' is a concept of the lower-case 'mike'—but this is not Mike's *self*-concept, as it ought to be.

Creary and Pollard (1985) have recently streamlined this theory and even have a mechanism (denoted: ↑I) for representing "the minimal concept of self"—a sort of first-person quasi-indicator: "This concept is the sense of the indexical pronoun *I*" (Creary & Pollard, 1985, p. 176). One limitation of their version of quasi-indicators is that it is restricted to the first person. More seriously, perhaps, is an apparent lack of a connection between ↑I and

[9] Again, (cf. n. 7), this assumes that '*S*' is a name.

its antecedent: If there are two candidates for an antecedent, how would the system be able to determine which is the correct one? The analysis of quasi-indicators to be presented below makes this link explicit. But at least Creary and Pollard are sensitive to the issues.

3.2.2.9. Maida and Shapiro. The research presented in this essay is a direct outgrowth of the work of Anthony S. Maida and Stuart C. Shapiro (1982). In particular, it began with an attempt to make a small correction to their analysis of 'John knows that he is taller than Bill' (Maida & Shapiro, 1982, p. 316)—once again, there was the problem of quasi-indicators. I shall discuss this aspect of their work in Section 4.5.4.1.

The thrust of Maida and Shapiro (1982) is to argue for a thoroughgoing intensionalism (see Section 4). I shall not rehearse their arguments here. Rather, I shall cite some of the principles that I have adopted from their work. The data base of an AI system capable of reasoning about the beliefs of itself and others must be such that:

- All items represented are intensional.
- Intensionally distinct items must have distinct representations.
- Distinct representations correspond to intensionally distinct items.

(These last two points can be referred to jointly as the Uniqueness Principle.)

- Intensionally distinct items that are extensionally the same are linked by a 'co-extensiveness' mechanism. (I have more to say about this in Rapaport [1985b].)

Maida (1985) has gone on to apply some of these techniques to reasoning about knowledge, though in a somewhat different style from the approach taken here and concentrating on somewhat different problems.

3.2.3. Conclusions. It should be clear from the above survey that there have been two chief problems with belief-representation systems. With scattered exceptions, there has been little attempt to seriously apply the insights of philosophers on the logic of belief; this will become clearer in what follows, as we look at some of the philosophical issues. And virtually none of the earlier systems[10] has shown how to deal with (much less shown any awareness of) the problem of quasi-indexical reference.

It is only fair to note that these criticisms are not necessarily fatal. What is important is that any theory dealing with beliefs that aspires to representational adequacy must have a way of handling quasi-indicators. In the next section, I show how this can be done.

---

[10] Moore (1980) and Creary and Pollard (1985) are the exceptions; but Moore's analysis only appeared in his 1980 dissertation and Creary and Pollard's work postdates the work presented here.

4. THE REPRESENTATION OF BELIEF

In this central section of the essay, we shall look briefly at: philosophical theories of, and computational needs for, intensionality; various distinctions between *de re* and *de dicto* beliefs; SNePS as a 'knowledge'-representation language; and, finally, a uniform and adequate representation, in SNePS, of *de re* and *de dicto* beliefs, nested beliefs, and quasi-indicators.

4.1. Intensionality

There has been much controversy in philosophy over the meaning of 'intension'. Although I shall not attempt to define it here, intensional approaches to problems in logic and language may be roughly characterized as ones that place more of an emphasis on "meanings," or mental states and events, than on truth values, denotations, or nonmental entities (things in the external world) (cf. Brody, 1967, p. 64, 67; Shapiro & Rapaport, in press.)

4.1.1. Intensional Contexts. Included among *intensional contexts* are such modalities as 'it is necessary that...', '*A* believes that...' '*A* is looking for...', and so on.

Among the intensional contexts are the *intentional* ones, those, such as belief contexts, that involve intentional—or psychological—verbs.[11] Typically, the objects of (psychological acts expressed by) intentional verbs need not exist or be true. Thus, one can think about unicorns or believe false propositions. Such objects are called 'intensional entities'.

Among the intentional verbs are those expressing *propositional attitudes:* 'believe', 'know', and so on. These express attitudes that a cognitive agent might take towards a proposition. 'Look for', 'think about', and so forth, are attitudes whose linguistic expression form intentional contexts, but they are not *propositional* attitudes.

Another kind of intentional context is provided by what can be called *practitional* attitudes (cf. Castañeda, 1975c): 'John ought (to)...', 'John intends (to)...', and so on. In 'John ought to sing', an "ought-to-do" operator applies to the "practition" 'John to sing'; in 'John wants Mary to sing', John has the attitude of wanting towards the practition 'Mary to sing', and in 'John intends to sing', he has the attitude of intending towards the practition 'he* to sing'.

4.1.2. Referential Opacity and Quantifying In. Two problems that must be faced by anyone dealing with intensional language are those of referential opacity and 'quantifying in.'

---

[11] This is, roughly, Brentano's sense of 'intentional', in which "intentionality" is a distinguishing mark of mental phenomena (cf. Brentano, 1874/1960; Chisholm, 1967; and Section 4.1.3.2.). By 'intentional' verbs, I do *not* mean such verbs as 'kick', as in 'John hated Lucy, so he (intentionally) kicked her'—that is, he kicked her *on purpose.*

A linguistic context is *referentially opaque* if substitution of coreferential constituents does not preserve truth value, and is *referentially transparent* otherwise.[12]

One must also take care when *quantifying into* intensional contexts. This can also be viewed as a substitutional issue: the replacement of one of the constituents by a bound variable. One's allegiance to intensional versus extensional approaches to logic, language, and ontology can often be determined by one's attitude towards one aspect of this problem: Suppose that Lucy, after having read a fantasy story, looks for a unicorn. Is there a unicorn that she is looking for? If you are inclined to answer "No" *solely on the grounds that unicorns do not exist,* then you might look askance at intensional contexts, because they cannot be thus *existentially* quantified into. But if you are inclined to answer "Yes" (or perhaps to hesitate), on the grounds that she is (or might be) looking for the (specific) unicorn that she read about, then you are willing to admit *intensional entities* into your ontology.

4.1.3. Intensional Theories in Philosophy. At the risk of oversimplifying, we might say that most contemporary theories of intensional entities can be traced back, through Alonzo Church (1951, 1973, 1974) and Rudolf Carnap (1956), to Gottlob Frege's (1892/1970c) essay, "On Sense and Reference," or else to Alexius Meinong's (1904/1971) essay, "The Theory of Objects."

4.1.3.1. The Fregean Approach. Frege distinguished between the *Sinn* (sense, meaning) and the *Bedeutung* (denotation, reference, referent, meaning) of words, phrases, and sentences. Each word or phrase *expresses* a sense, and each sense *determines* a referent. For instance, the sense expressed by 'the Morning Star' is, roughly, "the last starlike object visible in the morning sky"; this, in turn, determines a referent, namely, a certain astrophysical object, which is also called 'Venus'.[13] Frege introduced this distinction chiefly in order to eliminate it from his logical foundations for mathematics, which, he claimed, could be handled *without* recourse to senses; for mathematics, all that counted was the referents of sentences[14] and their constituents. The referent of a sentence is its *truth value,* the referent of a noun phrase is an *object,* and (arguably) the referent of a predicate is a *concept.*[15] The sense of a sentence is a *thought* (or *proposition*). However, since Frege wanted the referent of a sentence to be a function of its constituents, he had to complicate matters when it came to intensional sentences. The *referent* of a sentence that occurs within an intensional con-

---

[12] Alternatively, a context is referentially opaque if its extension (its truth value in the case of a sentence, its referent in the case of an NP, etc.) is not a function of the extensions of its parts.

[13] Note that the term 'Venus' expresses a *different* sense, but that sense determines the *same* referent as the one determined by the sense of 'the Morning Star'.

[14] Strictly: the referent determined by the sense expressed by the sentence!

[15] Cf. Frege 1891/1970a, pp. 30–31; 1982/1970b, p. 43, esp. fn.; Dummett 1967, p. 231. It is important to note that a Fregean "concept" is a technical notion (cf. Section 3.2.2.2. n. 6).

text is its ordinary *sense.* (But, he claimed, this complication did not arise in the case of mathematics.)

Various philosophers have attempted to formalize these notions (e.g., Church, 1951, 1973, 1974; Carnap, 1956; Montague, 1974) often using the essentially extensional techniques of possible worlds. I shall not go into these theories here, except to note that senses (and their more formal theoretical counterparts) are intensional entities. All words and phrases have senses, even 'unicorn', but not all have referents.

4.1.3.2. The Meinongian Approach. Alexius Meinong, a student of Brentano[16], took a more psychological approach. He analyzed psychological experiences into three components: a psychological *act* (e.g., such acts as believing, desiring, or thinking); an *object* of the act (e.g., that which is believed, or desired, or thought about); and a *content* of the act, which "directs" it to its object. (For details, see Meinong, 1904/1971; Findlay, 1963; Rapaport, 1976b, 1978, 1979.) All psychological acts are directed to objects; this is the Thesis of Intentionality, first suggested by Brentano (1874/1960) as a distinguishing mark of mental, as opposed to physical, phenomena. But not all objects exist, or, as Meinong humorously put it: "There are objects of which it is true that there are not such objects" (Meinong, 1904/1971, p. 490). That is, I can think about unicorns *in exactly the same way* that I can think about cats, and I can believe that $1 + 1 = 3$, even though the object of my belief is false (or does not exist, as Meinong would have put it).[17] The object of a propositional attitude is called an *objective;* the object of a thought is called an *objectum.* Like Fregean senses, Meinongian objectives and objecta are intensional entities.

Contemporary theories that are Meinongian in spirit include those of Castañeda (1972, 1975a, 1975b, 1975c, 1977, 1979; cf. Rapaport, 1976b, 1978), Terence Parsons (1980; cf. Rapaport, 1976b, 1978, 1985a), Richard Routley (1979; cf. Rapaport, 1984a), and Rapaport (1976b, 1978, 1979, 1981, 1982). All of these theories can handle referential opacity and quantifying in. Typically, substitution of coreferentials (more generally: of identicals) in intentional contexts is blocked on the grounds that the items to be substituted are intensionally *distinct.* And most cases of quantifying in are allowed, because quantifiers are presumed to range over *intensional entities.*[18]

4.2. The Need for Intensional Entities in "Knowledge" Representation

As has been forcefully argued by Woods (1975), Brachman (1977), Shapiro

---

[16] As was Edmund Husserl, the founder of phenomenology.

[17] Actually, he would have said it 'lacked being', in particular, it did not 'subsist'. For details and more motivation, see Rapaport (1976b, 1978).

[18] The sort of case that is not so easily handled is the one where, just because Lucy is looking for a unicorn, it does not follow that there is (even in the nonexistentially loaded sense of 'there is') a *particular* unicorn that she is looking for—any one will do. But this problem is beyond our present scope (cf. Rapaport, 1976b for more detail).

(1981), and Maida and Shapiro (1982), an AI system that is going to interact with a human, or that is intended to be a model (or simulation) of a human mind, must be able to represent and reason about intensional entities. And, as Maida and Shapiro have stressed, they *only* need to deal with intensional entities. I would also claim, in the Kantian spirit (cf. Section 2.5), that they *cannot* deal with *extensional* entities. What is lacking from such schemes is a full-blown theory of intensional entities. Any of the ones mentioned here would no doubt suffice, though I favor the Meinongian approach, because of its psychological underpinnings. (Castañeda's theory is especially appropriate for SNePS [cf. Rapaport, 1985b].)

4.3. *De Re* and *De Dicto*

In the philosophical literature, a belief *de dicto* is treated as a psychological act of belief whose object is a proposition—a '*dictum*'—and a belief *de re* is treated as a psychological act of belief whose object is (to use the Meinongian term) an objectum—a '*res*'.

For example, suppose that Ralph sees the person whom he knows to be the janitor stealing some government documents, and suppose—unknown to Ralph—that the janitor has just won the lottery. Then Ralph believes *de dicto* that the janitor is a spy, and he believes *de re* that the lottery winner is a spy. That is, if asked, Ralph would assent to the proposition 'The janitor is a spy'; but he merely believes *of the man* whom *we* know to be the lottery winner that he is a spy—Ralph would not assent to 'The lottery winner is a spy'.

Much of the philosophical literature on *de re* and *de dicto* belief concerns the relations between these, conceived as *ways* of believing. I do not believe that there are two such distinct psychological modes of belief. But be that as it may, it seems clear to me that there *are* two distinct ways of *reporting* an agent's beliefs, which may, for better or worse, be called *de re* and *de dicto*.[19] Thus, if what we are interested in is expressing or communicating the actual "content" of Ralph's belief, we would need to report it in a *de dicto* fashion. If, however, we are *not* concerned to, or *cannot,* communicate his belief in that manner, then we can (or must) report it in a *de re* fashion.

Traditionally viewed, a belief *de dicto* is a referentially opaque context, whereas a belief *de re* is referentially transparent. Thus, the inference

(10) Ralph believes [*de dicto*] that the janitor is a spy.
The janitor = the lottery winner.

Ralph believes [*de dicto*] that the lottery winner is a spy.

is invalid. Moreover, its conclusion not only presents *false* information, it represents a *loss* of information, namely, of the information about the propositional 'content' of Ralph's belief. On the other hand,

[19] Parsons (1980, p. 46, fn. 10) can be read as taking this interpretation.

(11) Ralph believes [*de re*] of the janitor that he is a spy.
The janitor = the lottery winner.

Ralph believes [*de re*] of the lottery winner that he is a spy.

is valid. But the conclusion conveys just as little information about Ralph's *actual* belief *de dicto* as does the first premise.

Castañeda (1970, p. 167ff) prefers to distinguish between *propositionally transparent* and *propositional opaque* constructions: The former display the internal make-up of the proposition; the latter don't. What I call a *de dicto belief report* is referentially opaque but propositionally transparent, whereas what I call a *de re belief report* is referentially transparent but propositionally opaque. It is the propositional kind of opacity and transparency that is important for communication and for representational issues in AI.

Ordinary language, however, does not distinguish these. That is, without an explicit device (e.g., as in (10) and (11) above), belief sentences can be interpreted as *either* de re *or* de dicto. This poses a pragmatic problem for the ultimate project. At this stage of investigation, however, I shall adopt the following conventions from Castañeda (1970, p. 174ff): Where A names or describes an agent and p a proposition, and x ranges over noun phrases,

(C1) Any sentence of the form

A believes that p

will be the canonical representation of a *de dicto* belief report.

(C2) Any sentence of the form

A believes of x that p

will be the canonical representation of a *de re* belief report, where x names or describes the objectum.

An alternative to (C2) is

(C3) x is believed by A to be F

where F names or describes the property predicated of x. Whereas (C3) has the advantage that the objectum is outside of the belief context, (C2) has the computational advantage that its grammatical structure is a generalization of (C1)'s structure, as well as the philosophical advantage that in both (C1) and (C2), the objective of the belief can be treated as being propositionally transparent (cf. Castañeda, 1970, p. 174ff).

Finally, a *de se* belief may be taken to be a belief about oneself (cf. Lewis, 1979, p. 521). Such a belief can be reported as being *de re* or *de dicto*. Consider the following:

(DD.DS) John believes that he* is rich.

(DR.DS) John believes of himself that he is rich.

The first will be taken here by convention as a *de dicto* report of John's *de se* belief; all such reports involve the use of quasi-indicators. The second

will be taken here by convention as a *de re* report of John's *de se* belief.[20] Both (DD.DS) and (DR.DS) are mutually consistent: for (DR.DS) might be true because John might believe that the editor of *Cognitive Science* is rich, yet not believe that he* is the editor of *Cognitive Science*. We, who know that he *is* the editor, can report his belief about the editor by the *de re/de se* sentence (DR.DS). On the other hand, (DD.DS) is *in*consistent with John's believing that he* is not rich.

4.4. The SNePS Semantic Network Processing System

SNePS (Shapiro, 1979; Shapiro & Rapaport, in press) is a facility for building semantic networks that represent propositions (objectives) and individuals, properties, and relations (objecta), and for retrieving and deducing information from these networks. A SNePS network consists of nodes linked by labeled, directed arcs. The nodes and arcs have the following features, among many others:

(S.1) Each constant node represents a unique concept.

(S.2) Each concept represented in the network is represented by a unique node.

(S.3) Arcs represent nonconceptual, binary relations.

(S.4) Deduction rules are propositions, and so are represented by nodes.

In the context of the sort of system considered here, *nondominated* nodes—that is, nodes with no arcs pointing to them—represent *beliefs of the system*. *All* nodes, whether dominated or not, are in the 'mind' of the system. To use Meinongian terminology, all nodes represent Meinongian objects of the *system's* psychological acts: Dominated nodes represent objecta; nondominated nodes represent objectives.[21]

As a simple example, the sentence

(12) John is rich.

could be represented in SNePS by the network of Figure 1. The OBJECT-PROPERTY case frame is used to represent simple subject-predicate propositions. So, the nondominated node m5 represents the system's belief that something (viz., whatever is represented by node m3) has the property represented by m4. The LEX arc points from a node to a tag used by the ATN parser-generator to attach a word to the node. (The node labeled 'rich', however, could also be considered to be the system's concept of the *word* 'rich'. Cf. Maida and Shapiro [1982, p. 303] and Shapiro and Rapaport [in press] for details on the semantic interpretation of the LEX arc.) The OBJECT-

[20] Because of the use of 'himself', it might be hard to 'hear' (DR.DS) as *de re*. It is to be understood as 'John believes of the person whom *we* know to be him that he (that person) is rich'. But this is too long and grammatically complex to be a canonical representation.

[21] The complete set of all possible nodes—whether actually in the network or not—corresponds neatly to Meinong's notion of *Aussersein* (cf. Rapaport, 1976b, 1978, 1985b).

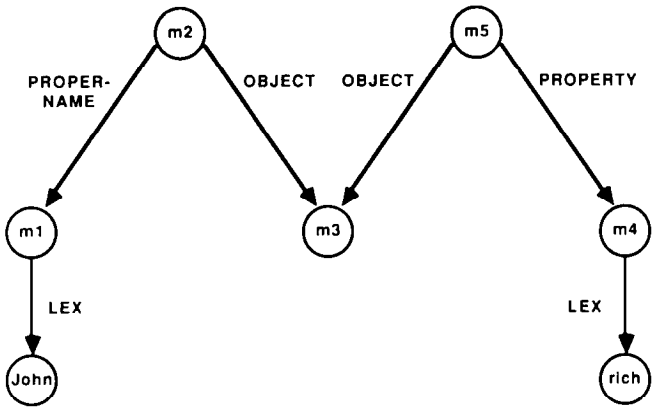


Figure 1. SNePS network for 'John is rich'.

PROPER-NAME case frame is used to identify objecta by name. So, the nondominated node m2 represents the system's belief that something (viz., whatever is represented by node m3) is named by the name whose concept is m1. More idiomatically, (12) has been analyzed as

(12A) Something named 'John' is rich.

or, more perspicuously,

(12B) $\exists x$[Proper-Name(x, 'John') & Rich(x)]

or, more accurately,[22]

(12C) ($\exists u, v, x, y, z$)[OBJECT(y, x) & PROPER-NAME(y, u) & LEX(u, 'John') & OBJECT(z, x) & PROPERTY(z, v) & LEX(v, 'rich')]

where the quantifier ranges over objecta in the system's belief space, and the predicates are represented by the arcs. (In particular, u = m1, x = m3, y = m2, z = m5, and v = m4.)

There is nothing sacrosanct about this analysis. We could just as well have represented (12) by the network of Figure 2. Here, m7 represents the system's belief that the entity represented by m3 is a person, and m9 represents the system's belief that the entity represented by m3 is male. Clearly, the amount of detail one wishes to build into the network analysis depends

[22] The predicate logic formulas here and elsewhere are to be taken only as suggested ways of reading the SNePS networks, which are the fundamental representation. In particular, no semantic interpretation of the predicate logic formulas is being offered, though any such interpretation would probably take the variables to range over intensional objects (and the variables in the scope of the second argument place of 'Believes' [see (13A)] to range over intensional objects in the belief space of the entity in its first argument place), in order to be consistent with the interpretation of the networks. (Cf. Shapiro and Rapaport [in press] for details of the semantic interpretation.)

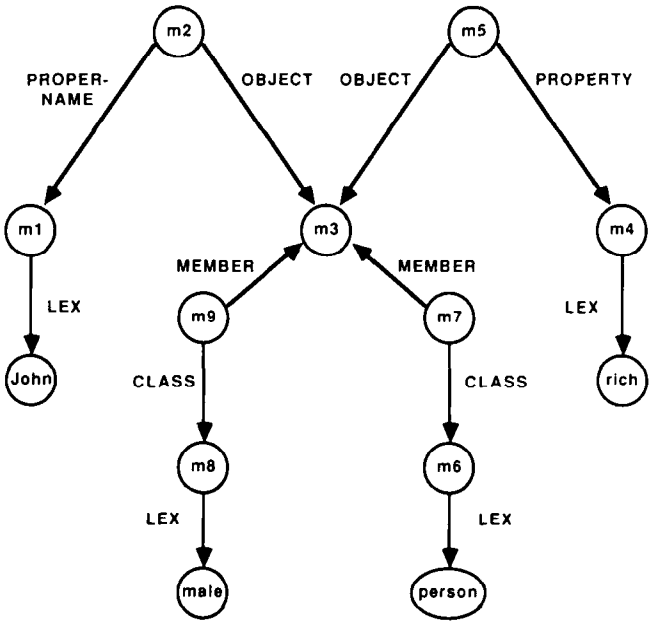


Figure 2. Another SNePS network for 'John is rich'.

on one's ontological and linguistic theories. Rather than worry about such details here, I shall use the analysis of Figure 1, because it contains the essential information for my purposes.

To represent the system's beliefs about the beliefs of others, an AGENT-ACT-OBJECT case frame is used. Thus,

John believes that *p*.

will be represented by the network fragment of Figure 3. Here, m6 represents the system's belief that the person whom the system believes to be named 'John' believes m5, where m5—in a concrete case—will be a node representing, *roughly,* the system's *de dicto* report of John's belief. I say 'roughly', because, in fact, the system can never represent John's belief *exactly;* it can only represent John's belief using its *own* concepts. I shall return to this point shortly (cf. Section 2.4 and Sections 4.5.2, 4.5.3), but it should be noted that humans have precisely the same limitations.

4.5. Belief Representation in SNePS

I can now be a bit more precise. However, to make things easier for the reader, I shall not present a general scheme for representation, but only representations of actual—though representative—sentences.

4.5.1. De Re and *De Dicto* Beliefs.

The *de dicto* belief report

(13) John believes that Lucy is sweet.

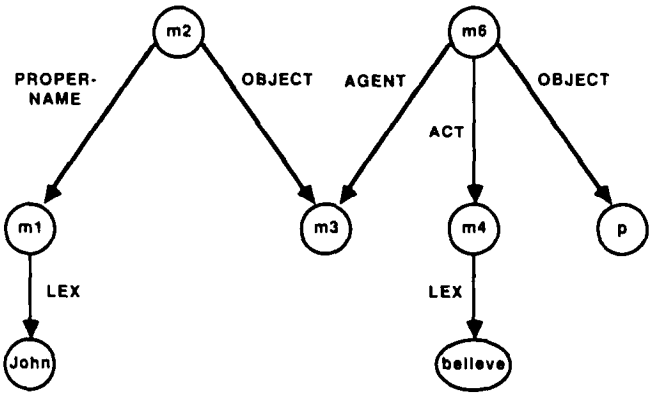


Figure 3. Fragment of SNePS network for 'John believes that p'.

will be represented by the network of Figure 4. A predicate-logic reading of (13) might be:

(13A) $\exists x$[Proper-Name(x, 'John') & Believes(x, $\exists y$[Proper-Name(y, 'Lucy') & Sweet(y)]]

That is, the system believes three things: that someone is named 'John', that he (John) believes that someone is named 'Lucy', and that he (John) believes that she (i.e., the person he believes to be named 'Lucy') is sweet.

To simplify the graphical notation, which will quickly become complex, Figure 4 can also be drawn as in Figure 5. The idea here is that, since m8 and m11 are both nodes in an AGENT-ACT-OBJECT case frame with the same AGENT and the same ACT, we can eliminate the redundant arcs from the *graphical representation* of the network by using the box notation. *This is a notational convenience only.*

A *de re* belief report,

(14) John believes of Lucy that she is sweet.

will be represented by the network of Figure 6. A predicate-logic reading of (14) might be:

(14A) ($\exists x$, y)[Proper-Name(x, 'John') & Proper-Name(y, 'Lucy') & Believes(x, Sweet(y))]

That is, the system believes three things: that someone is named 'John', that someone is named 'Lucy', and that he (John) believes of her (Lucy) that she is sweet. Note that here the system has no beliefs about how John represents Lucy.[23]

[23] These analysis, although arrived at independently, bear a structural similarity to analyses by Castañeda (1970, p. 176ff) and by Chisholm (1976, p. 13).

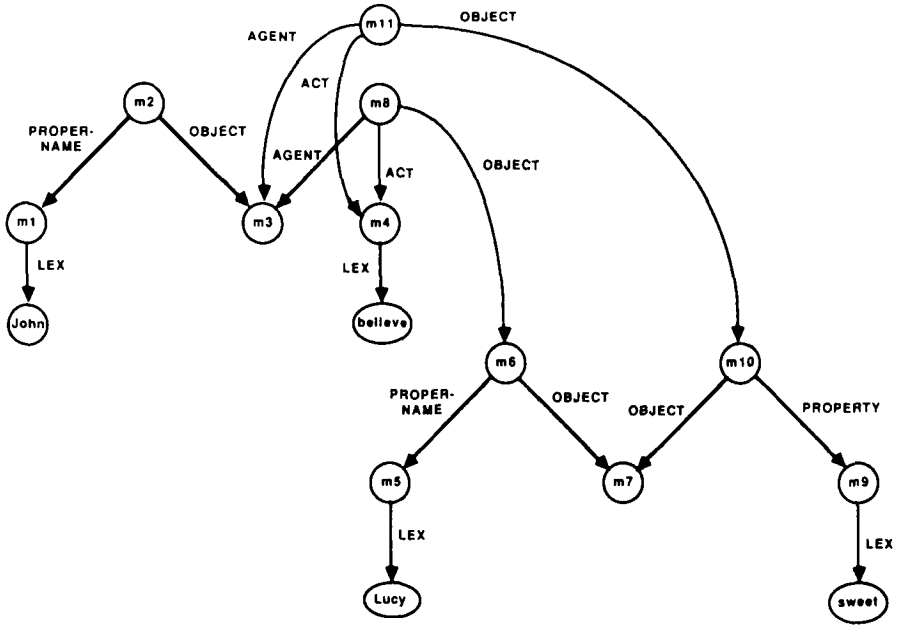


Figure 4. A SNePS network for the *de dicto* belief report 'John believes that Lucy is sweet'.

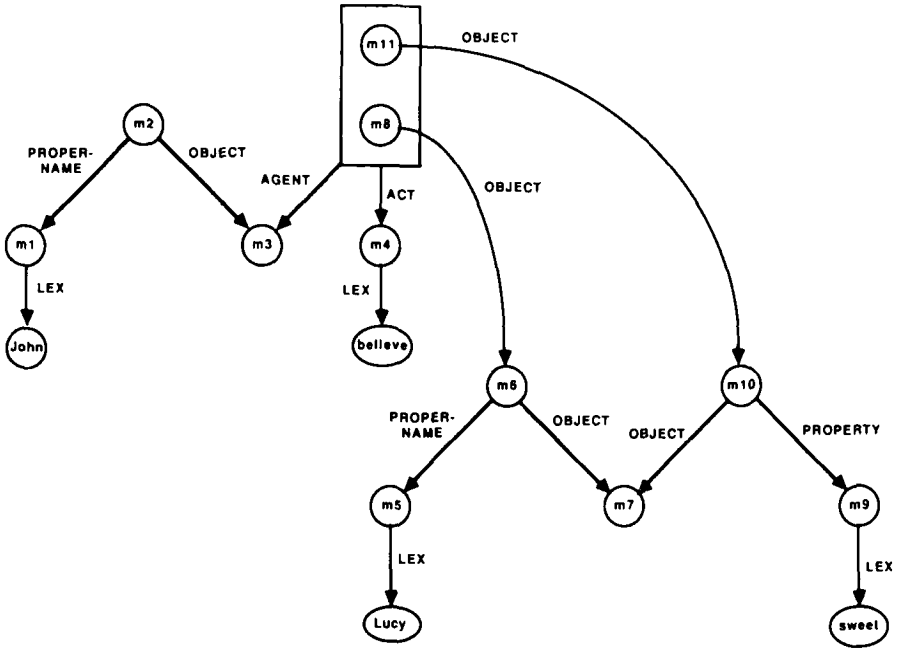


Figure 5. An alternative graphical representation of the network in Figure 4.

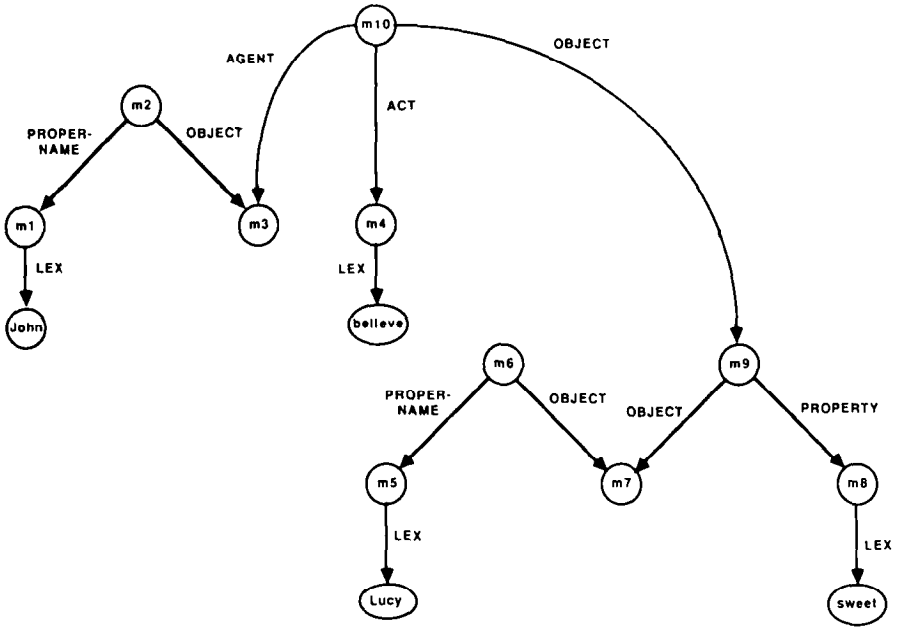


Figure 6. A SNePS representation of the *de re* belief report 'John believes of Lucy that she is sweet'.

We can also combine these. But here we must be careful. In the absence of prior knowledge of coextensiveness, the entities *within* a belief context should be represented *separately* from entities that might be coextensive with them but that are *outside* the context. Suppose that the system's beliefs include that a person named 'Lucy' is young and that John believes that a (possibly different) person named 'Lucy' is rich. This is represented by the network of Figure 7. The section of network dominated by nodes m10 and m14 is the system's *de dicto* representation of John's belief. That is, m14 is the *system's* representation *of* a belief that *John* would express by 'Lucy is rich', and it is represented *as* one of John's beliefs. Such nodes are considered as being in the system's representation of John's *belief space*. If it is later determined that the "two" Lucies are the same, then a node of coextensiveness would be added, as in Figure 8 (node m16). (Cf. Maida and Shapiro [1982, pp. 303–304] and Rapaport [1985b] for discussions of the semantics of the EQUIV case frame.)

4.5.2. Nested Beliefs. The representational scheme presented here can also handle sentences involving nested belief contexts. Consider

(15) Bill believes that Stu believes that Hector is philosophical.

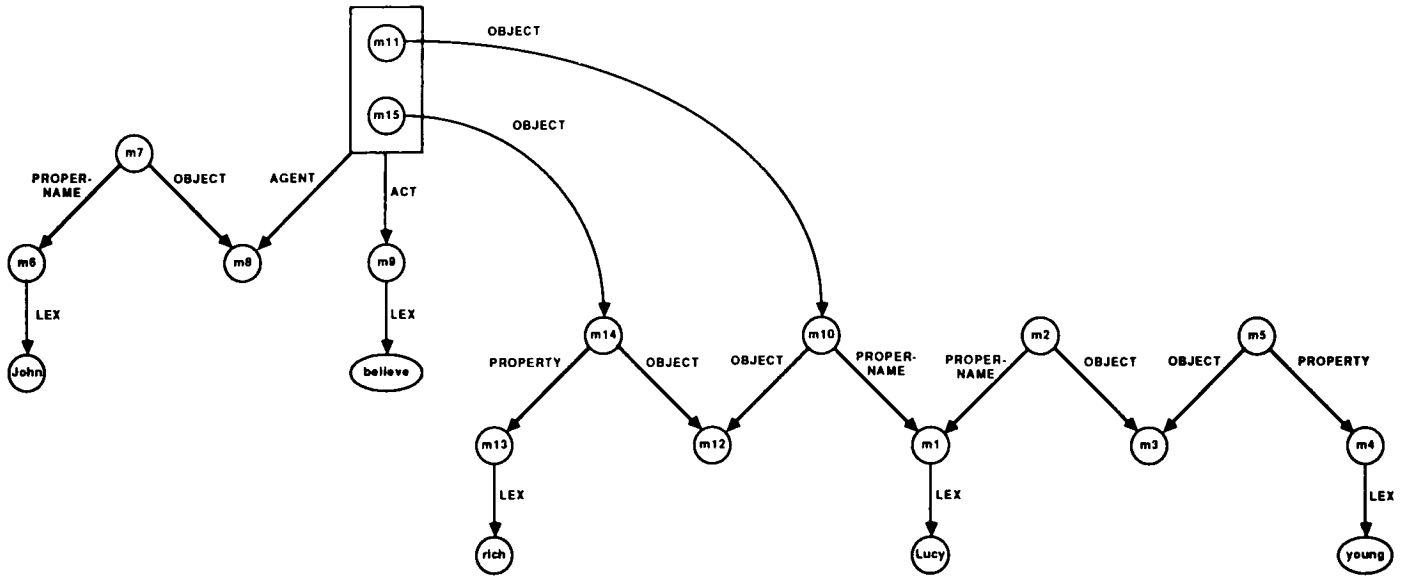


Figure 7. SNePS network representing that Lucy is young and that John believes that someone named 'Lucy' is rich.

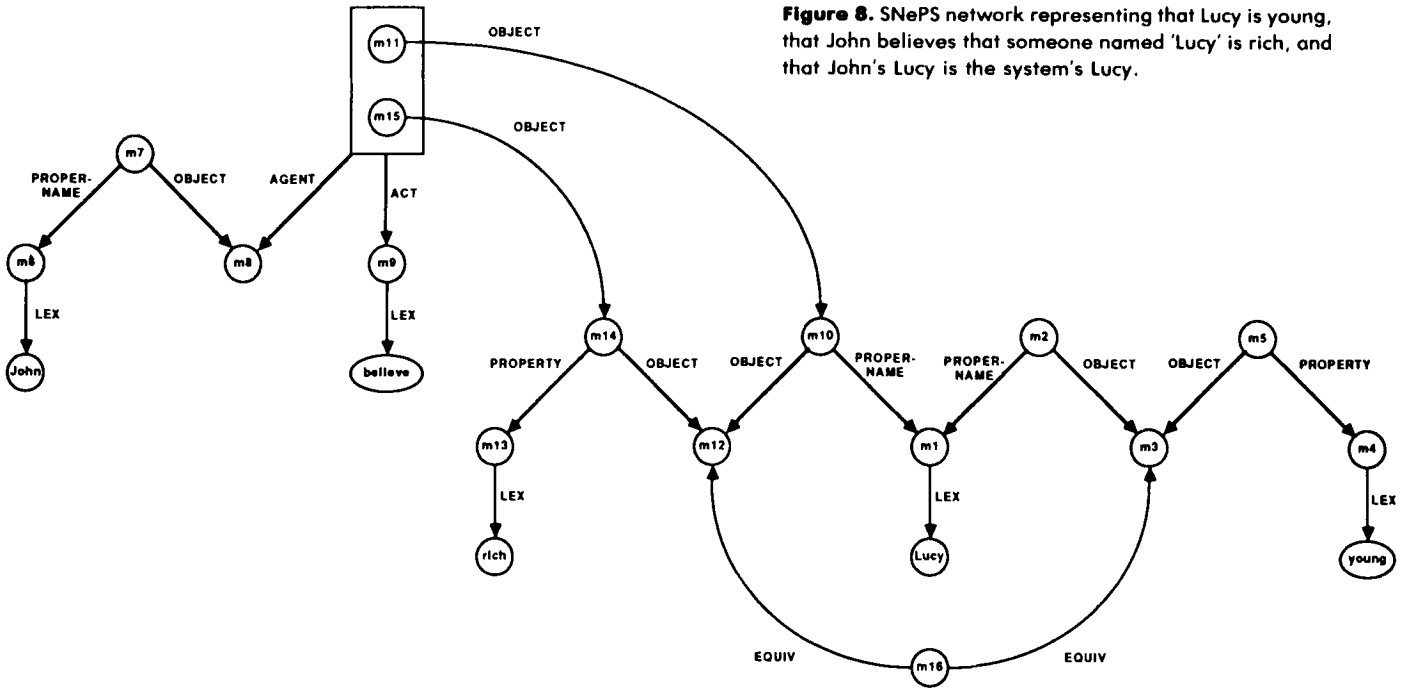


Figure 8. SNePS network representing that Lucy is young, that John believes that someone named 'Lucy' is rich, and that John's Lucy is the system's Lucy.

The interpretation of this that I am most interested in representing treats (15) as the system's *de dicto* representation of Bill's *de dicto* representation of Stu's *de dicto* belief that Hector is philosophical. On this interpretation, we need to represent the *system's* Bill, the system's representation of *Bill's* Stu, and the system's representation of Bill's representation of *Stu's* Hector. These can be characterized by stacks implemented as LISP lists: (Bill system), (Stu Bill system), and (Hector Stu Bill system), respectively.[24] Sentence (15) is represented by the network of Figure 9.

In the implementation, such a network is built recursively as follows: The parser maintains a stack of "believers"; the top element on the stack is the current believer, the bottom element is the system. Each time a belief sentence is parsed, it is made the object of a belief of the previous believer in the stack. Structures are shared wherever possible. This is accomplished by use of a new SNePS User Language (Shapiro, 1979) function, *forb-in-context* (find-or-build in a belief context), which, given a description of a node and a belief context (a stack of believers), either finds the node in that context if it exists there (i.e., if it is already in the believer's belief space), or else builds it in that context (i.e., represents the fact that it is now in the believer's belief space). Thus,

Bill believes that Stu believes that Hector is Guatemalan.

would modify the network of Figure 9 by adding new beliefs to (Bill system)'s belief space and to (Stu Bill system)'s belief space, but would use the same nodes to represent Bill, Stu, and Hector.

4.5.3. Some Comments. Before turning to the representation of quasi-indicators, it is worth discussing two representational choices that could have been made differently.

First, in Figures 4-6, I am assuming that the system's concept of sweetness (Figures 4-5, node m9; Figure 6, node m8) is also the system's concept of (Lucy system)'s concept of sweetness. This assumption seems warranted, because *all* nodes are in the system's belief space. If the system had reason to believe that *its* concept of sweetness differed from Lucy's, this could—and would have to—be represented (cf. Section 6).

Second, there are, *prima facie,* at least two other possible case frames to represent situations where an agent has several beliefs (of which the *de dicto* case is merely one instance). For example, instead of the network of Figure 4 or 5, we could have used either of the networks of Figures 10 and 11.

But the case frame for node m10 in Figure 10 makes no sense linguistically: Syntactically, a belief sentence is an SVO sentence (or, it has an AGENT-

[24] Each item on the list (except the first) can be considered as a subscript of the preceding item.

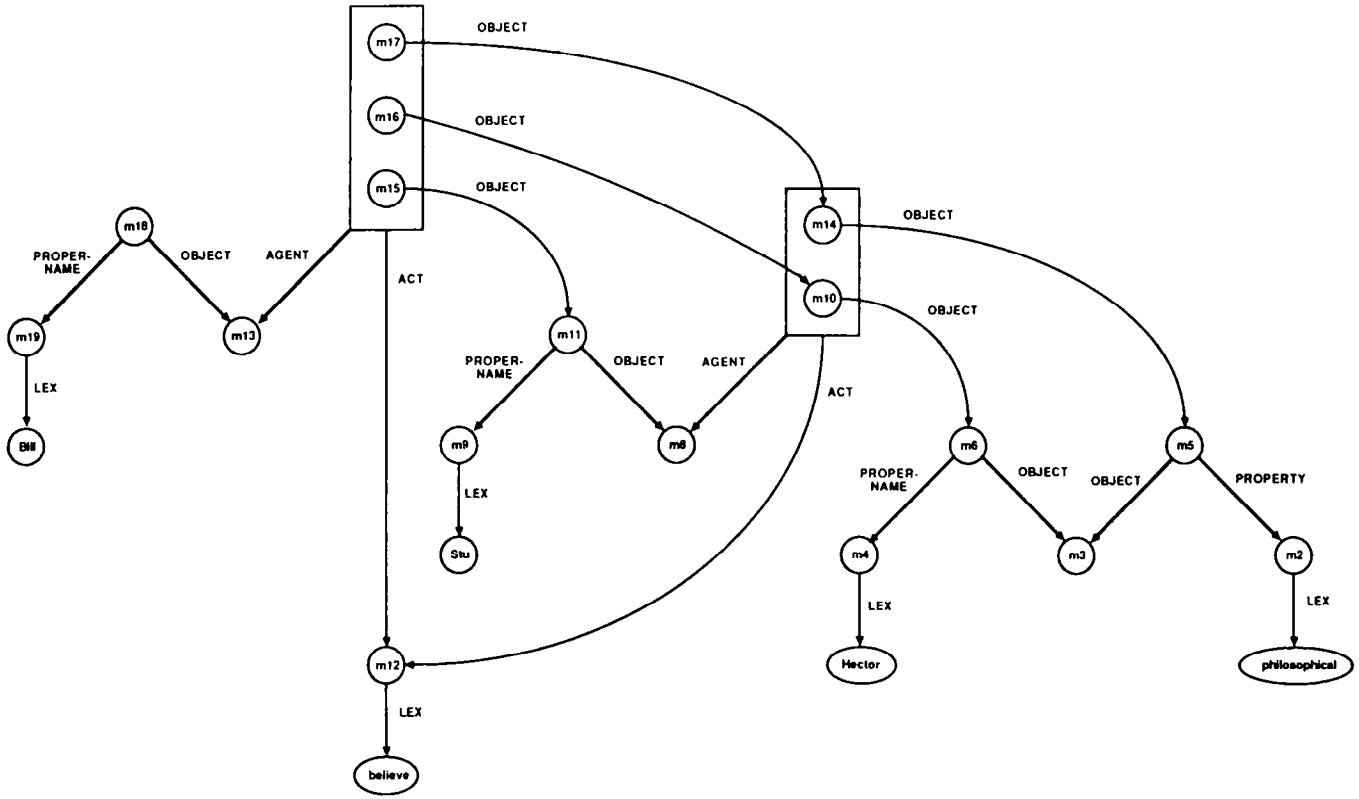


Figure 9. SNePS network for 'Bill believes that Stu believes Hector is philosophical .

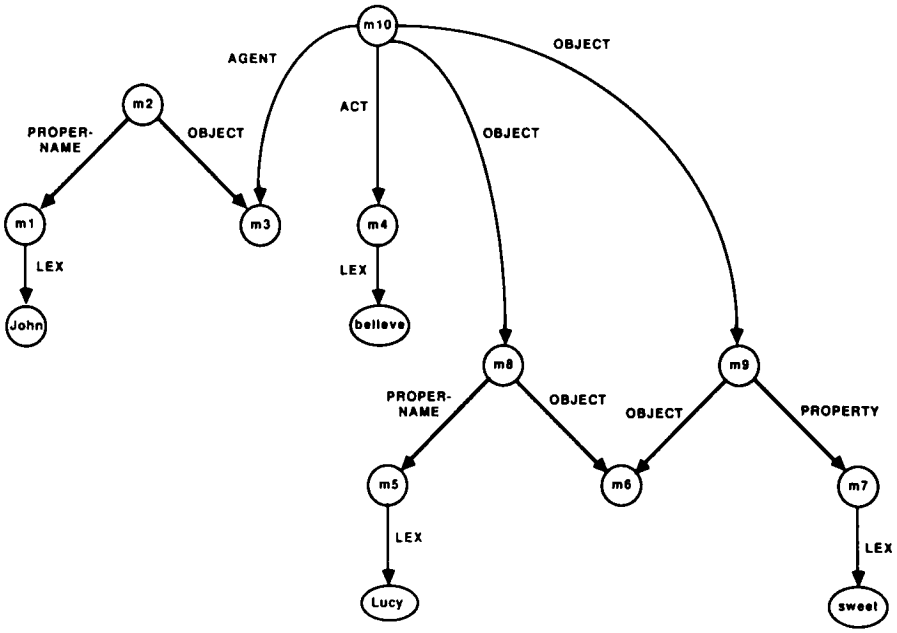


Figure 10. Alternative SNePS network for 'John believes that Lucy is sweet'.

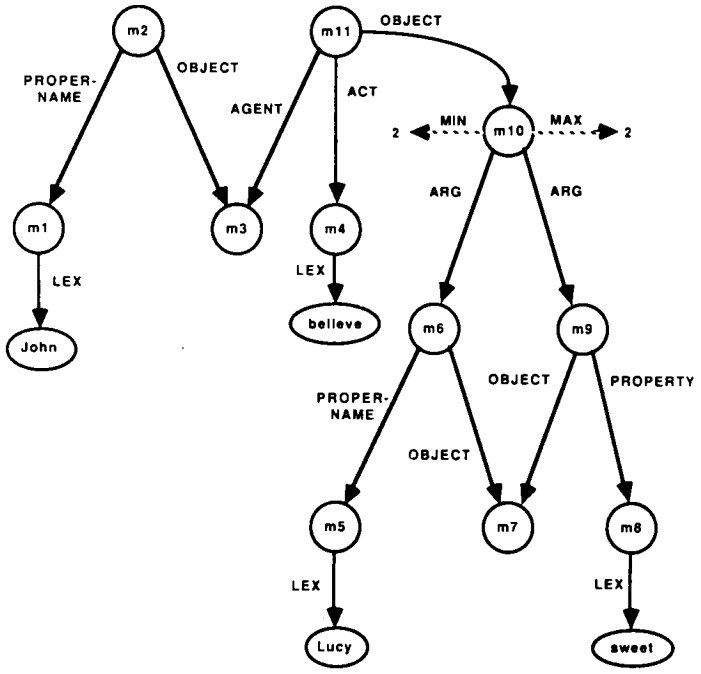


Figure 11. Another alternative SNePS network for 'John believes that Lucy is sweet'.

ACT-OBJECT case frame); it is not an SVOO, SVOOO, and so on, sentence. Having a single OBJECT arc to a 'supernode', in the manner of Hendrix (1979, p. 64) might be better, but still odd linguistically. In addition, as new beliefs are added, the case frame would have to change. A related disadvantage to this alternative[25] is that m10 is a *single* proposition with two objects (i.e., there are two Meinongian objectives of the act of believing), rather than two propositions each with a single object (i.e., two acts of believing, each with a single objective). Thus, a question like "Who believes *p*?" or an addition to the database that *one* of the reports about John's beliefs is false would apply to *all* of the beliefs, not just to one of them.

In Figure 11, node m10 is the SNePS representation of the conjunction of m6 and m9. One objection to this format is that the nodes in *John's* belief space are explicitly conjoined, unlike the nodes in the *system's* belief space; there seems no good reason for this structural dissimilarity. Moreover, retrieval of the information that John believes that someone is sweet would be difficult, since that information is not explicitly represented in Figure 11.

4.5.4. De Se Beliefs.

4.5.4.1. The Representation. To adequately represent *de dicto* reports of *de se* beliefs, we need the strategy of separating entities in different belief spaces (see Section 4.5.1). Consider the possible representation of

(2) John believes that he* is rich.

shown in Figure 12 (adapted from Maida & Shapiro, 1982, p. 316).

This suffers from three major problems. First, it is ambiguous: It could conceivably be the representation of (2) as well as

(16) John believes that John is rich.

But, as we have seen, (2) and (16) express quite different propositions; thus, they should be separate items in the data base.

However, Figure 12 cannot represent (16). For then we would have no easy or uniform way to represent (2) in the case where John does not know that he is named 'John': Figure 12 says that the person (m3) who is named 'John' and who believes m6, believes that that person is rich: and this would be false in the amnesia case.

But Figure 12 cannot represent (2) either, for it does not adequately represent the quasi-indexical nature of the 'he' in (2): Node m3 represents both 'John' and 'he', hence is both inside and outside the intentional context, contrary to both of the properties of quasi-indicators discussed in Section 3.1.

---

[25] This disadvantage was pointed out to me by Stuart C. Shapiro.

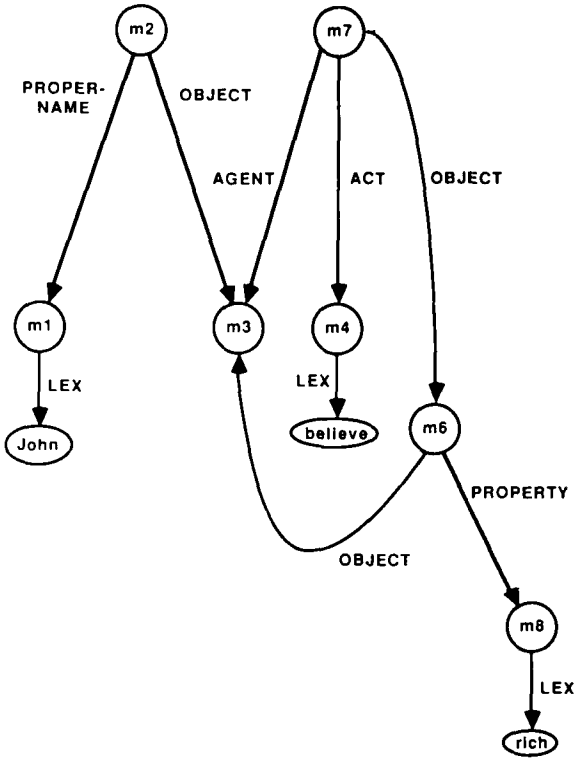


Figure 12. A possible SNePS network for 'John believes that he* is rich'.

Finally, because of these representational inadequacies, the system invalidly 'infers' (16) from (2)–(17):

(2) John believes that he is rich.
(17) He = John

(16) John believes that John is rich.

simply because premise (2) is represented by the same network as conclusion (16). (I shall return to this in Section 5.1.)

Rather, the general pattern for representing such sentences is illustrated in Figure 13. The role that "he*" plays in the English sentence is represented by node m5; its quasi-indexical nature is represented by means of node m7.

That nodes m3 and m5 must be distinct follows from the separation principle. But, because m5 is the system's representation of John's representation of himself, it must be within the system's representation of John's belief space; this is accomplished via nodes m7 and m6, representing John's belief that m5 is his 'self-representation'. Node m6, with its EGO arc to m5, represents, roughly, the proposition 'm5 is me' (cf. Section 4.5.4.2.).

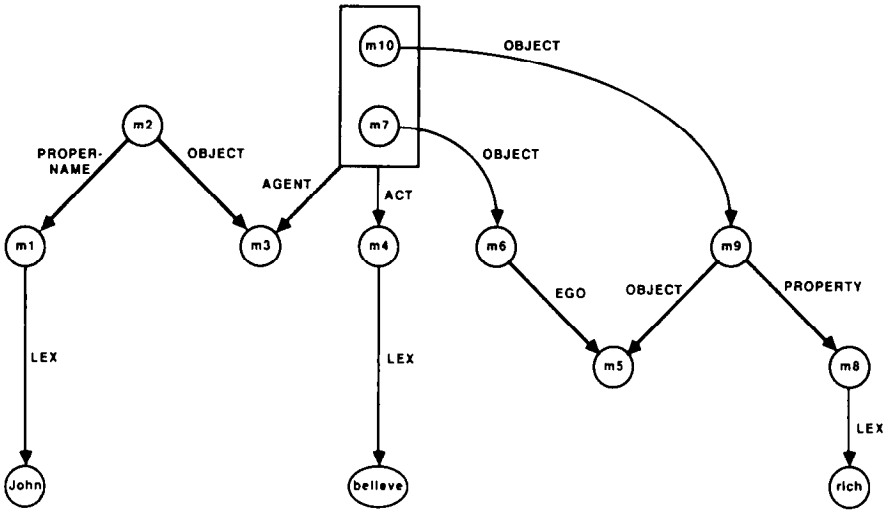


Figure 13. SNePS network for 'John believes that he* is rich'. (Node m5 is the system's representation of John's 'self-concept', expressed by John as 'I' and by the system as 'he*'.)

This representation of quasi-indexical *de se* sentences is thus a special case of the general schema for *de dicto* representation of belief sentences. When a *de se* sentence is interpreted *de re,* it does not contain quasi-indicators, and can be handled by the general schema for *de re* representations. Thus, the *de re* report

(18) John believes of himself that he is rich

(or: John is believed by himself to be rich) *would* be represented by the Maida-Shapiro network of Figure 12.

As a final example, consider the following:

(3) John believes that the editor of *Cognitive Science* is rich.
(4) John believes that he* is not rich.
(19) John is the editor of *Cognitive Science.*
(18) John believes of himself that he is rich.

If John is unaware of the truth of (19), then the system ought to be able to infer—and, hence, to represent—(18). We can represent the data base resulting from the input of this sequence of information and inference by the network of Figure 14. Here, nodes m11 and m7 represent (4) (node m10 represents a SNePS negation of node m9); nodes m15 and m17 represent (3); m18 represents (19); and m20 represents (18).

4.5.4.2. The EGO Arc. One representational issue requires discussion: the EGO arc. The node pointed to by the EGO arc is the system's represen-

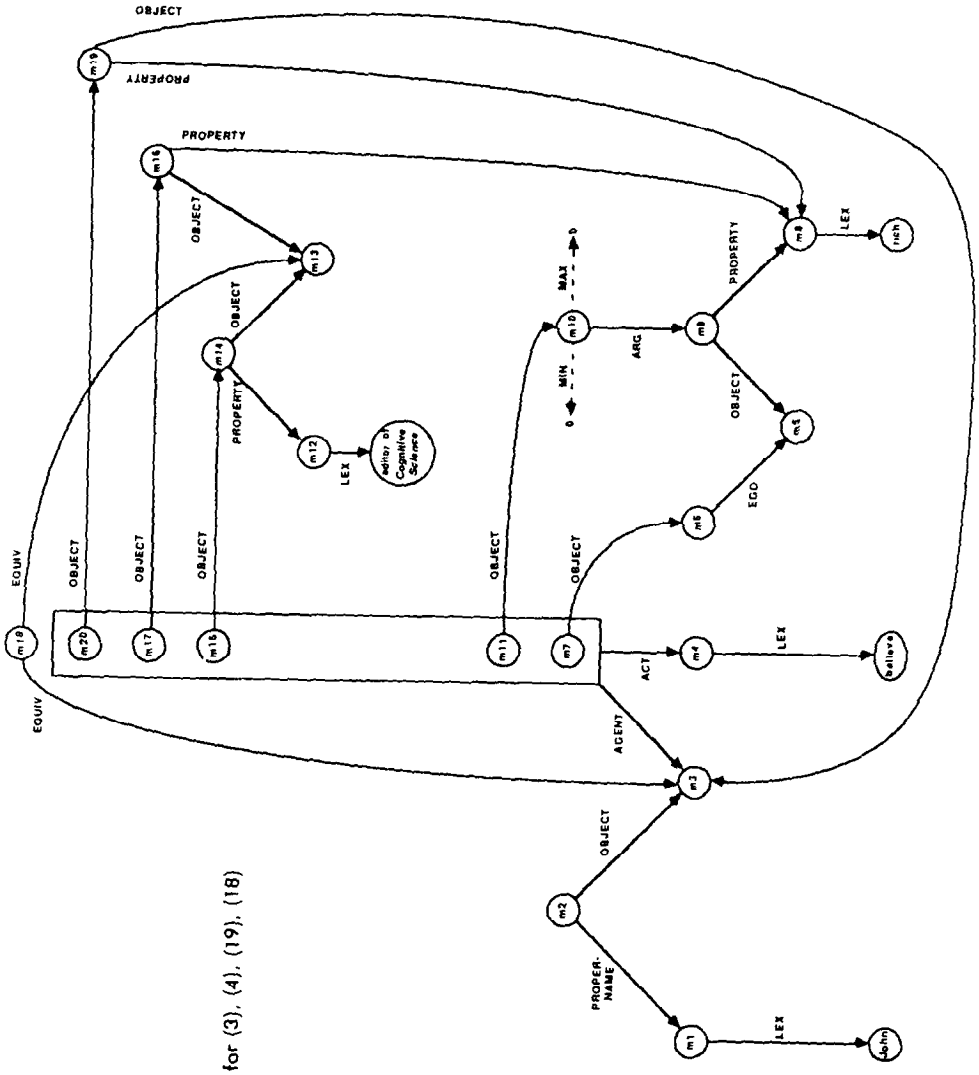


Figure 14. SNePS network for (3), (4), (19), (18)

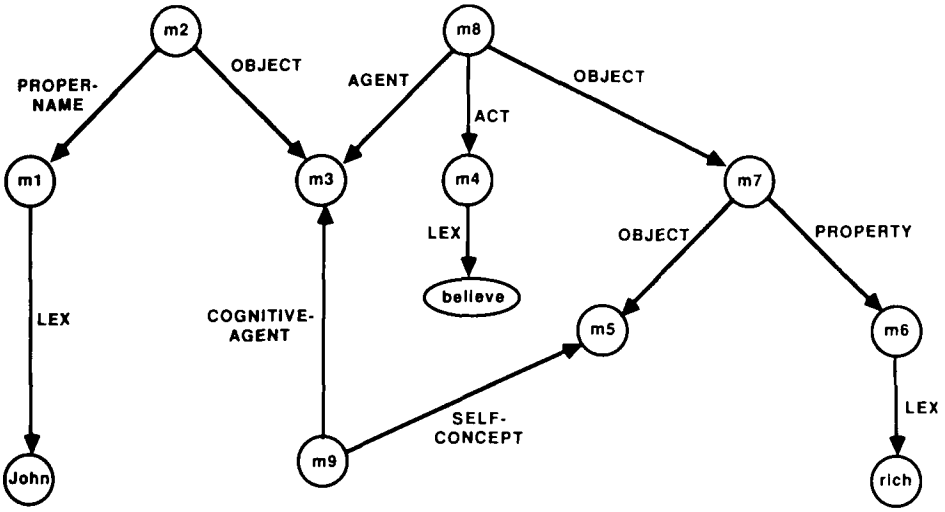


Figure 15. Alternative SNePS network for 'John believes that he* is rich'.

tation of what might be called John's *self-concept,* that is, John's 'model' of himself (cf. Minsky, 1968). The EGO arc provides the link with the quasi-indicator's antecedent (cf. the discussion of Creary and Pollack [1985] in Section 3.2.2.8).

Again, let us consider alternative representations. The network of Figure 15, which is the simplest alternative, uses a COGNITIVE-AGENT/SELF-CONCEPT case frame to "identify" m5 as being m3's (John's) self-concept (a kind of 'merger' of m5 with m3). But m5 needs to be *inside John's* belief space (to be on a par with Figure 13), which this network does not do. Note that, in Figure 13, node m7 also links John (m3) with his self-concept (m5)[26], as does Figure 15's m9, but it does so while placing m5 within John's belief space.

Another alternative is the network of Figure 16. Here, we do have m5 within John's belief space, but we also have John there—*as well as* outside it—which violates the principle of the separation of belief spaces. And, again, any viable role played by such a case frame can also be played by node m7 of Figure 13.

Finally, consider the network of Figure 17. Here, the idea is to have a case frame analogous to the PROPER-NAME–OBJECT one, which might be called the QI-EGO case frame. Node m7 represents the proposition

[26] By means of the path from (i.e., the relative product of) the converse-EGO arc from Figure 11's m5 to m6, through the converse-OBJECT arc from m6 to m7, to the AGENT arc to m3.

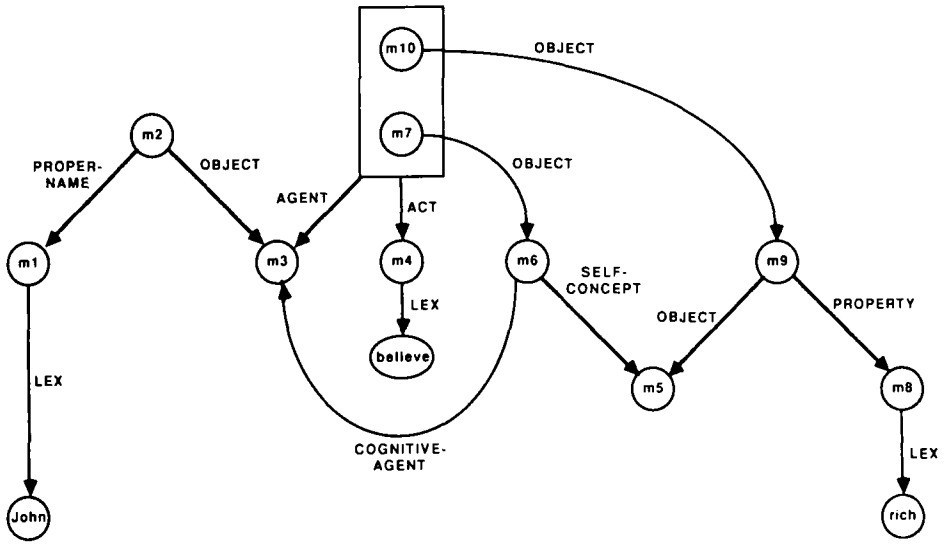


Figure 16. Another alternative SNePS network for 'John believes that he* is rich'.

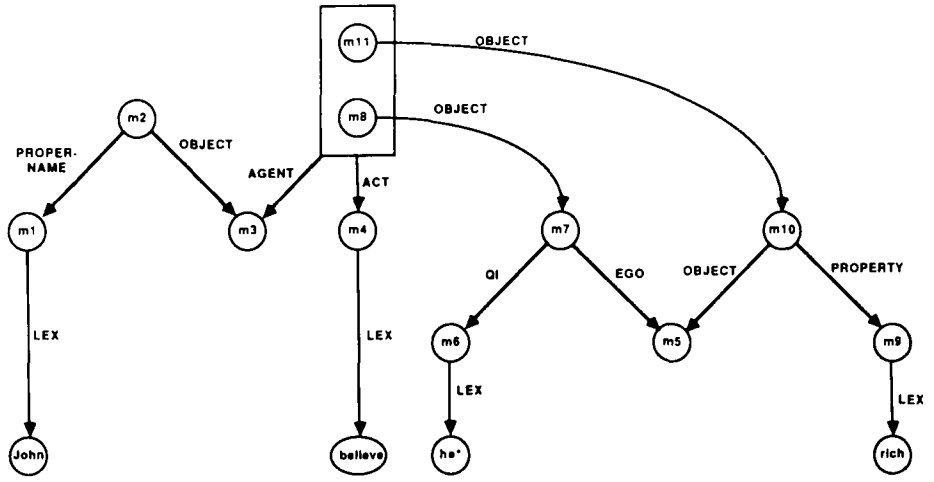


Figure 17. A third alternative SNePS network for 'John believes that he* is rich'.

(believed by John) that m5 is he himself. But, of course, that is *not* the proposition believed by John. Rather, he believes a proposition that *he* would express as 'm5 is me'. And that is precisely what the original EGO arc of Figure 13 is intended to capture.[27]

5. INFERENCES

5.1. Inferences Using the Maida and Shapiro Network

Recall the problem with the Maida and Shapiro network for 'John believes that he* is rich' (Section 4.5.4.1.). It was ambiguous in the sense that a parser taking English sentences as input and producing (SNePS) networks as output would produce the network of Figure 12 as the parse of *both* of the following sentences:

(2) John believes that he* is rich.

(16) John believes that John is rich.

But, as we have seen, each should be parsed differently: For if John does not believe that he* is John, then it might be the case that (16) is true while it is *not* the case that (2) is true, or vice versa. Thus, the system would 'infer' (16) from (2) together with

(17) He is John.

But the 'inference' would be *static* rather than *dynamic;* that is, rather than having to use an inference rule to *build a new piece of network* corresponding to (16), the system would merely *retrieve* (2) ("find" it, in SNePS terminology), because there would only *be one* net for (2) *and* (16).

This, it is to be emphasized, is a problem with Maida and Shapiro's *representation,* not their rules. Indeed, the point is precisely that no rules are involved here. Nevertheless, given the insensitivity of their networks to quasi-indexical reference, the actual rules they present are more powerful than they should be. The observations that follow hold for any representa-

---

[27] At least two philosophers have suggested analyses of *de dicto/de se* beliefs that are structurally similar to the lone EGO arc. Chisholm would say that such a belief report conveys the information that John has an 'individual essence' m5 and that he believes that whatever has m5 is rich (cf. Chisholm, 1977, p. 169). And Perry introduces an *ego function* that maps a person to that person's "special sense"; a Perry-style analysis of our *de dicto/de se* belief would be:

$\exists s[s = \text{ego(John)} \;\&\; \text{Believes(John, Rich}(s))]$

(cf. Perry, 1983, p. 19, 25). There are difficulties with both of these suggestions, which I shall not go into here (cf. e.g., Castañeda, 1983, p. 326f), and there are also more complicated cases that need to be examined. But these are topics for future investigation. It should, perhaps, be mentioned that Castañeda has indicated (in conversation) that my representation accurately captures his theory.

tions that do not recognize the need for quasi-indicators, including those of Cohen and Perrault (1979) and Clark and Marshall (1981) (cf. Section 3).

For instance, consider Maida and Shapiro's Rule 2 and one of its proffered instances (paraphrased from Maida & Shapiro, 1982, p. 330):

(Rule 2) $(\forall x, y, z, R)$[Believes(x, Rxy) & Believes(x, Equiv(y, z)) → Believes(x, Rxz)]

(P) If John believes that Jim's wife is Sally's mother and *he* (John) believes that *he** wants to meet Jim's wife, then *he* (John) believes that *he** wants to meet Sally's mother (and this is so regardless of whether Jim's wife *in fact is* Sally's mother).

[italics and use of 'he*' added]

The problem is that (P) is *not* an instance of Rule 2, because the 'x' in 'Rxy' cannot be replaced by a quasi-indicator whose antecedent is the 'x' in the first argument place of 'Believes'; it can only be replaced by the *same* value that 'x' gets. Thus, the following *would* be a legitimate instance of Rule 2:

(P1) If John believes that Jim's wife is Sally's mother and he (John) believes that *John* wants to meet Jim's wife, then he (John) believes that *John* wants to meet Sally's mother.

There can be no legitimate use of Rule 2 to sanction (P). For if there were, it would also sanction the following:

(P2) If John believes that Jim's wife is Sally's mother and he (John) believes that *he** wants to meet Jim's wife, then he (John) believes that *John* wants to meet Sally's mother.

(P3) If *he** believes that Jim's wife is Sally's mother and *he** believes that *he** wants to meet Jim's wife, then *he** believes that *he** wants to meet Sally's mother.

But these, as we have seen, should not be sanctioned: (P2) should not be, because it might not be true; and neither should (P3), because quasi-indicators cannot appear outside the 'believes-that' context.

5.2. The Maida and Shapiro Approach to Inference

However, Maida and Shapiro take an *approach* to rules of inference for sentences involving propositional attitudes that is quite appropriate to the distinction between *propositional* opacity and transparency (discussed in Section 4.3), which emphasizes the priority of *communication:*

> A system that conforms to the Uniqueness Principle [cf. Section 3.2.2.9.] does not need the substitutivity of equals for equals as a basic reasoning rule, because no two distinct nodes are equal. Co-referentiality between two nodes must be asserted by a proposition. It requires inference rules to propagate assertions from one node to another node which is co-referential with it. Thus, *intensional representation implies that referential opacity is the norm* [italics added] and transparency must be explicitly sanctioned by an inference process. (Maida & Shapiro, 1982, p. 300)

Referential opacity is the norm. Or, to use Castañeda's terminology, *propositional transparency is the norm.* Inferences such as those just discussed *should* be blocked, because, as noted, they represent a *loss* of information.

5.3. Adequate Inference Rules

What inference rules do we want? For one thing, we want some rule to sanction (P): The following will do this:[28]

(R1) $(\forall x, y, z, w, \mathrm{R})[\mathrm{Believes}(x, \mathrm{R}wy) \,\&\, \mathrm{Believes}(x, \mathrm{Equiv}(y, z)) \rightarrow \mathrm{Believes}(x, \mathrm{R}wz)]$

The difference between Rule 2 and (R1) is the extra variable, w, which is needed in order to make (R1) a rule of inference that *preserves propositional transparency.* Both (P) and (P1) are instances of (R1), but (P2) and (P3) are not. ((P3) is not, because the quasi-indicator 'he*' cannot be a value of x.)

Similarly, if we want the following inference to be valid:[29]

(3) John believes that the editor of *Cognitive Science* is rich.
(20) John believes that he* is the editor of *Cognitive Science.*

(2) John believes that he* is rich.

as well as the inference with (2) and (3) switched, we would need the following propositionally transparent rule:

(R2) $(\forall x, y, z, \mathrm{F})[\mathrm{Believes}(x, \mathrm{F}y) \,\&\, \mathrm{Believes}(x, z = y) \rightarrow \mathrm{Believes}(x, \mathrm{F}z)]$

To sanction inferences involving referentially transparent transitive predicates, further rules are needed. Consider the following valid inferences:

(21) (i) John is taller than Bill.
(ii) Bill is the editor of *Cognitive Science.*

(iii) John is taller than the editor of *Cognitive Science.*

(22) The editor of *Cognitive Science* is rich.
(19) John is the editor of *Cognitive Science.*

(12) John is rich.

(23) John wrote a memo to the editor of *Cognitive Science.*
(19) John is the editor of *Cognitive Science.*

(24) John wrote a memo to (a) John.
(b) himself.

[28] Note that what is ultimately needed is a rule *schema,* because, in the general case, R will vary in arity.

[29] James Moor (personal communication, 1984) is skeptical of this desire: For it is perfectly possible for John to have the beliefs expressed by (3) and (20), but fail to draw the proper inference. Clearly, I am using an idealized notion of belief here. The issue Moor points out is related to the sorts of issues discussed by Konolige, and is at the performance level.

(Note that 'wrote a memo to' might be considered to have a referentially opaque reading: here, it is being used referentially transparently.)

To sanction (21), for instance, we need an additional premise to the effect that *being taller than* is referentially transparent:

RefTransp(tall)

(this could be part of the lexical entry for 'tall'), and we need a rule for EQUIV for relations:

(R3) $(\forall x, y, z, R)[\text{RefTransp}(R) \& Rxy \& \text{Equiv}(y, z) \rightarrow Rxz]$

The other inferences would be handled similarly.

It should be noted that in SNePS, (R1)–(R3) would not be second-order rules (even though the predicate-logic formulations I have given here quantify over predicates), because the relation is represented by a node, not an arc. Hence, 'R' in (R3) is never in predicate position and, hence, can be quantified over without ascending to second-order logic. (Cf. Shapiro [1979, pp. 192–193] which offers a third notation, and SNePS feature (S.4), mentioned in Section 4.4, as well as the discussion in Rapaport [1985b, p. 44].)

6. BELIEF REVISION

Belief-revision systems have typically been concerned with the problem of revising a system's data base in the light of new information. This is not simply a matter of adding new information, because the new information might be inconsistent with the old information (cf. Doyle, 1979). If the system is to be capable of reasoning about the beliefs of others, or if several users might contribute information (new beliefs) to the system's data base, the system must be able to keep track of the source of each belief. (Cf. Martins [1983] for an interesting approach using a form of relevance logic.) Because belief-revision systems must be capable of representing and reasoning about beliefs, they must be sensitive to the logical and intensional research issues. (We are currently investigating the mutual applicability of the system presented here with that of Martins [1983].)

The system under consideration here *is* capable of handling sequences of new information that might require it to revise its beliefs by *node merging:* the process of "identifying" two nodes.[10] For instance, suppose that the system is given the following information at three successive times:

at time t1: (Lucy system) believes that (Lucy Lucy system) is sweet.
at time t2: (Lucy system) is sweet.
at time t3: (Lucy system) = (Lucy Lucy system).

[10] The material in this section is based on conversations with Shapiro.

Then it should build the networks of Figures 18–20, successively.

At time t1 (Figure 18), node m3 represents (Lucy system) and m7 represents (Lucy Lucy system).

At time t2 (Figure 19), m13 is built, representing the system's belief that (Lucy system) (who is not yet believed to be—and, indeed, might *not* be—(Lucy Lucy system)) is sweet (cf. Section 4.5.3).

At time t3 (Figure 20), m14 is built, representing the system's new belief that there is really only one Lucy. This is a merging of the two "Lucy" nodes. From now on, all properties of (Lucy system) will be inherited by the (Lucy Lucy system), by means of an inference rule for the EQUIV case frame (similar to rule (R3); cf. Maida & Shapiro, 1982, p. 330)—but not vice versa, because properties that (Lucy system) believes (Lucy Lucy system) to have are not necessarily properties that (Lucy system) is believed by the system to have.

7. CURRENT RESEARCH

7.1. The EGO Arc

The EGO arc has further applications in the representation of self-knowledge. For instance, it can be used to represent the system's belief that it* is (or is not!) a computer, as in Figure 21. At least one philosopher has claimed that "Machines lack an irreducible first-person perspective.... Therefore, machines are not agents" (Baker, 1981, p. 157). The implications of the EGO arc for being able to represent the system's first-person perspective are excit-

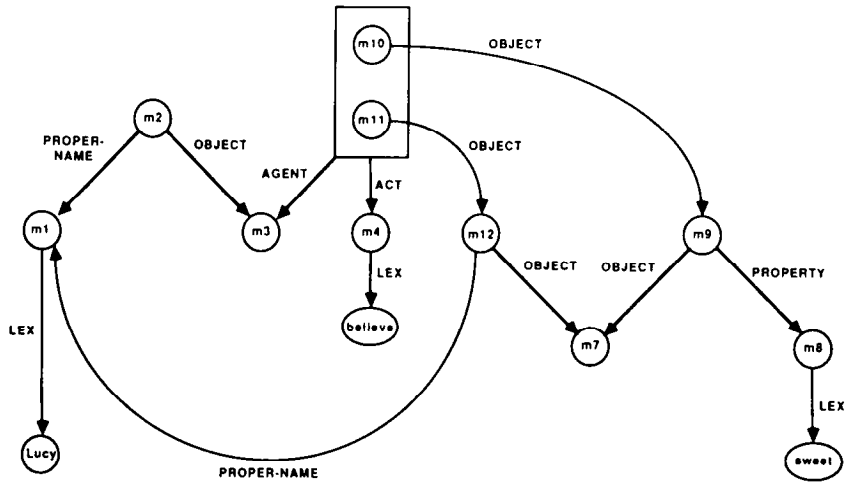


Figure 18. SNePS network at time t1, representing 'Lucy believes that Lucy is sweet'.

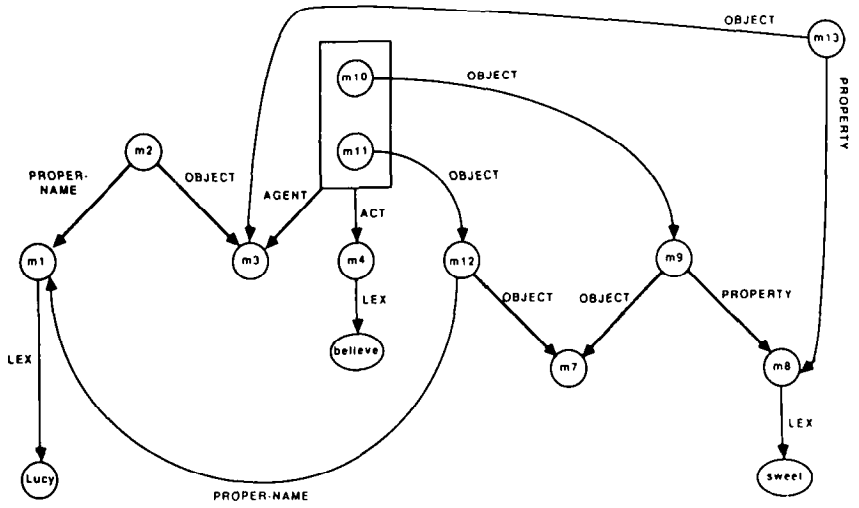


Figure 19. The SNePS network at time t2, modified from the network of Fig. 18, representing: 'Lucy believes that Lucy is sweet', and that Lucy (the believer) is sweet.

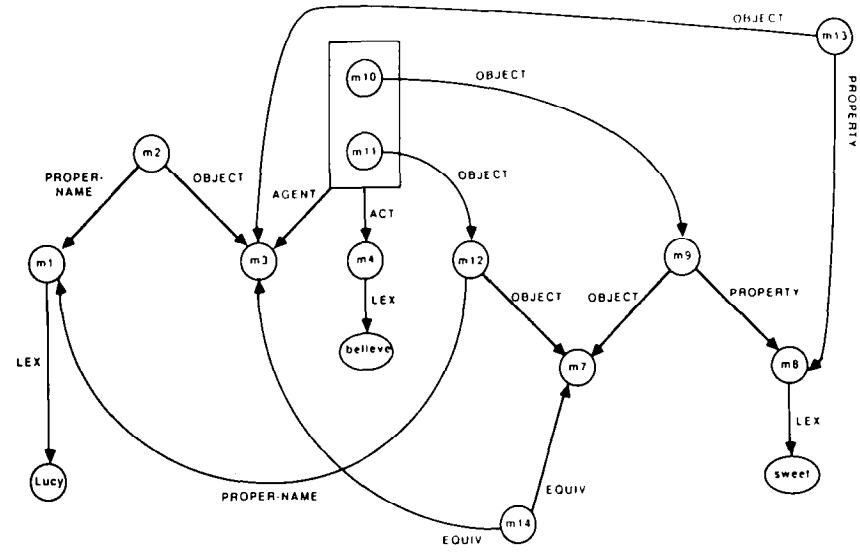


Figure 20. The SNePS network at time t3, modified from the network of Fig. 19, representing: 'Lucy believes that Lucy is sweet', that Lucy (the believer) is sweet, and that the system's Lucy is Lucy's Lucy.

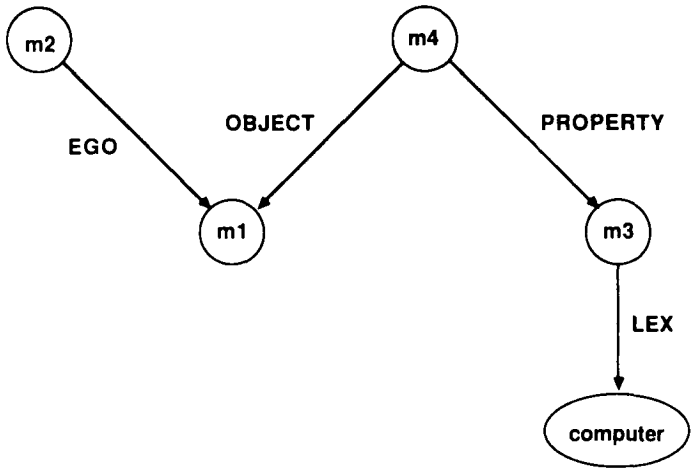


Figure 21. SNePS network for the system's belief that it* is a computer.

ing and worth exploring, especially in light of Castañeda's quasi-indexical theory of the first-person (Castañeda, 1968b).

7.2. Disambiguating Belief Reports

Rather than treating all sentences of the form

(25) *A* believes that *x* is F.

as canonically *de dicto* and all sentences of the form

A believes of *x* that F*x*.

as canonically *de re,* the system needs to be more flexible. In ordinary conversation, both sentences can be understood in either way, depending on context, including prior beliefs as well as idiosyncracies of particular predicates. For instance, given

(3) John believes that the editor of *Cognitive Science* is rich.

and the fact that John is the editor of *Cognitive Science,* most people would infer

(2) John believes that he is rich.

But given[31]

John believes that all identical twins are conceited.
Unknown to John, he is an identical twin.

[31] I owe this example to Carol Glass.

most people would *not* infer

John believes that he* is conceited.

Thus, we want to allow the system to make the most 'reasonable' or psychologically plausible (*de re* vs. *de dicto*) interpretations of users' belief reports, based on prior beliefs and on subject matter, and to modify its initial representation as more information is received.

Ordinarily, a user's belief report would be expressed using form (25). Currently; we are considering techniques for disambiguating belief reports on the basis of whether the system 'knows who' x is: Roughly, if the system believes that A knows who x is, then it should treat (25) as a *de re* report, or else it should treat it as a *de dicto* report, though things are not as simple as this. Somewhat more complicated situations arise during the dynamics of discourse that yield mixed *de re/de dicto* reports, which we are tentatively grouping under the name 'intermediate' (Wiebe & Rapaport, 1986). Inferencing using the system's belief space would then be used to determine whether the system believes that A knows who x is. In particular, we are investigating the analysis of 'knowing who' due to Boër and Lycan (1976, 1986). According to their theory, 'knowing who' is context-sensitive: We can only determine whether A knows who x is *for some purpose.*

7.3. Deictic Centers in Narrative

In joint work with other members of the SUNY at Buffalo Graduate Group in Cognitive Science, the system described here is being applied to the development of a model of a cognitive agent's comprehension of narrative text. The system embodying this model will represent the agent's beliefs about the objects, relations, and events in a narrative as a function of the form and context of the successive sentences encountered in reading it. In particular, we are concentrating on the role that spatial, temporal, and focal-character information plays in the agent's comprehension.

We propose to test the hypothesis that the construction and modification of a *deictic center* (cf. Fillmore, 1975) is of crucial importance for much comprehension of narrative. We see the deictic center as the locus in conceptual space-time of the objects and events depicted or described by the sentences currently being perceived. At any point in the narrative, the cognitive agent's attention is focused on particular characters (and other objects) standing in particular spatial and temporal relations to each other. Moreover, the agent 'looks' at the narrative from the perspective of a particular character, spatial location, or temporal location. Thus, the deictic center consists of a WHERE point, a WHEN point, and a WHO point. In addition, reference to character's beliefs, personalities, and so on, are also constrained by the deictic center.

'Knowing who' is of importance for computing the deictic center. In computing the deictic center, the system must be able to determine the current

values of the WHO, WHEN, and WHERE points. In particular, it must be able to determine *who* the current focal character is—the character that the system's attention is drawn to by the narrator. In some cases, the system's beliefs about the WHERE and WHEN points will help determine this; in others, the system's beliefs about who the focal character is will help determine the WHERE and WHEN points. The focal character is the character who is 'brought along' by shifts in the deictic center: If WHERE or WHEN change, so might WHO, and vice versa. Therefore, for the system to know who the focal character is requires knowledge of the rest of the deictic center, among other things.

Unfortunately, Boër and Lycan's analysis of 'knowing who' does not spell out the details of what a purpose is nor how it should be used in determining who someone is. Our project, however, provides a clear candidate for a purpose, namely, constructing the deictic center in order to comprehend the narrative. The system needs to be able to determine who the focal character is for the purpose of updating the deictic center (and not, say, for the purpose of providing a literary analysis of the narrative). The investigations of the linguists and psychologists in the research group will provide data for how this purpose will be used.

REFERENCES

Adams, R.M., & Castaneda, H.-N. (1983). Knowledge and belief: A correspondence. In J.E. Tomberlin (Ed.), *Agent, language, and the structure of the world.* Indianapolis, IN: Hackett.

Baker, L.R. (1981). Why computers can't act. *American Philosophical Quarterly, 18,* 157–163.

Barnden, J.A. (1983). Intensions as such: An outline. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence (IJCAI-83)* (p. 280–286). Los Altos, CA: Morgan Kaufmann.

Boër, S.E., & Lycan, W.G. (1976). Knowing who. *Philosophical Studies, 28,* 299–344.

Boër, S.E., & Lycan, W.G. (1980). Who, me?. *Philosophical Review, 89,* 427–466.

Boër, S.E., & Lycan, W.G. (1986). *Knowing who.* Cambridge, MA: MIT Press.

Brachman, R.J. (1977). What's in a concept: Structural foundations for semantic networks. *International Journal for Man-Machine Studies, 9,* 127–152.

Brand, M. (1984). *Intending and acting.* Cambridge, MA: MIT Press.

Brentano, F. (1960). The distinction between mental and physical phenomena. (D.B. Terrell, Trans.). In R.M. Chisholm (Ed.), *Realism and the background of phenomenology.* New York: Free Press. (Original work published 1874.)

Brody, B.A. (1967). Glossary of logical terms. In P. Edwards (Ed.), *Encyclopedia of philosophy* (Vol. 5). New York: Macmillan and Free Press.

Carberry, M.S. (1984). Understanding pragmatically ill-formed input. *Proceedings of the International Conference on Computational Linguistics (COLING-84)* (pp. 200–206). Morristown, NJ: Association for Computational Linguistics.

Carnap, R. (1956). *Meaning and necessity* (2nd ed.). Chicago: University of Chicago Press.

Castañeda, H.-N. (1966). 'He': A study in the logic of self-consciousness. *Ratio, 8,* 130–157.

Castañeda, H.-N. (1967a). On the logic of self-knowledge. *Noûs, 1,* 9–21.

Castañeda, H.-N. (1967b). Indicators and quasi-indicators. *American Philosophical Quarterly, 4,* 85–100.

Castañeda, H.-N. (1968a). On the logic of attributions of self-knowledge to others. *Journal of Philosophy, 54,* 439–456.

Castañeda, H.-N. (1968b). On the phenomeno-logic of the I. *Proceedings of the 14th International Congress of Philosophy.* Vienna: Herder.

Castañeda, H.-N. (1970). On the philosophical foundations of the theory of communication: Reference. *Midwest Studies in Philosophy, 2* (1977), 165–186.

Castañeda, H.-N. (1972). Thinking and the Structure of the World. *Philosophia, 4,* 3–40. (Reprinted in 1975 in *Critica, 6* (1972), 43–86.

Castañeda, H.-N. (1975a). Individuals and non-identity: A new look. *American Philosophical Quarterly, 12,* 131–140.

Castañeda, H.-N. (1975b). Identity and sameness. *Philosophia, 5,* 121–150.

Castañeda, H.-N. (1975c). *Thinking and doing.* Dordrecht: D. Reidel.

Castañeda, H.-N. (1977). Perception, belief, and the structure of physical objects and consciousness. *Synthese, 35,* 285–351.

Castañeda, H.-N. (1979). Fiction and reality: Their basic connections. *Poetica, 8,* 31–62.

Castañeda, H.-N. (1980). Reference, reality, and perceptual fields. *Proceedings and Addresses of the American Philosophical Association, 53,* 763–823.

Castañeda, H.-N. (1983). Reply to John Perry: Meaning, belief, and reference. In J.E. Tomberlin (Ed.), *Agent, language, and the structure of the world.* Indianapolis, IN: Hackett.

Chisholm, R.M. (1967). Intentionality. In P. Edwards (Ed.), *Encyclopedia of philosophy* (Vol. 4). New York: Macmillan and Free Press.

Chisholm, R.M. (1976). Knowledge and belief: 'De dicto' and 'de re'. *Philosophical Studies, 29,* 1–20.

Chisholm, R.M. (1977). Thought and its reference. *American Philosophical Quarterly, 14,* 167–172.

Church, A. (1950). On Carnap's analysis of statements of assertion and belief. In L. Linsky (Ed.), *Reference and modality.* London: Oxford University Press (1971).

Church, A. (1951). A formulation of the logic of sense and denotation. In P. Henle, H. Kallen, & S. Langer (Eds.), *Structure, method, and meaning.* New York: Liberal Arts Press.

Church, A. (1973). Outline of a revised formulation of the logic of sense and denotation. Part I, *Noûs, 7,* 24–33.

Church, A. (1974). Outline of a revised formulation of the logic of sense and denotation, Part II. *Noûs, 8,* 135–156.

Clark, H.H., & Marshall, C.R. (1981). Definite reference and mutual knowledge. In A.K. Joshi, B.L. Webber, & I.A. Sag (Eds.), *Elements of discourse understanding.* Cambridge: Cambridge University Press.

Cohen, P.R., & Perrault, C.R. (1979). Elements of a plan-based theory of speech acts. In B.L. Webber & N.J. Nilsson (Eds)., *Readings in artificial intelligence.* Palo Alto, CA: Tioga, (1981).

Creary, L.G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Proceedings of the Sixth International Joint Conference on Artificial Intelligence (IJCAI-79)* (Vol. I, pp. 176–181). Los Altos, CA: Morgan Kaufmann.

Creary, L.G., & Pollard, C.J. (1985). A computational semantics for natural language. *Proceedings of the Association for Computational Linguistics, 23,* 172–179.

Cresswell, M.J. (1980). Quotational theories of propositional attitudes. *Journal of Philosophical Logic, 9,* 17–40.

Dennett, D.C. (1978). Artificial intelligence as philosophy and as psychology. In D.C. Dennett (Ed.), *Brainstorms: Philosophical essays on mind and psychology.* Montgomery, VT: Bradford Books.

Dennett, D.C. (1983). Intentional systems in cognitive ethology: The 'Panglossian Paradigm' defended. *Brain and Behavioral Sciences, 6,* 343–390.

Doyle, J. (1979). A truth maintenance system. *Artificial Intelligence, 12,* 231–272.

Dummett, M. (1967). Gottlob Frege. In P. Edwards (Ed.), *Encyclopedia of philosophy* (Vol. 3). New York: Macmillan and Free Press.

Feldman, R.H. (1977). Belief and inscriptions. *Philosophical Studies, 32,* 349-353.

Fillmore, C. (1975). *Santa Cruz lectures on deixis.* Bloomington, IN: Indiana University Linguistics Club.

Findlay, J.N. (1963). *Meinong's theory of objects and values.* (2nd ed.). Oxford: Clarendon Press.

Frege, G. (1970a). Function and concept (P.T. Geach, Trans.). In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege.* Oxford: Basil Blackwell. (Original work published 1891)

Frege, G. (1970b). On concept and object (P.T. Geach, Trans.). In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege.* Oxford: Basil Blackwell. (Original work published 1892)

Frege, G. (1970c). On sense and reference. (M. Black, Trans.). In P. Geach & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege.* Oxford: Basil Blackwell. (Original work published 1892)

Hardwig, J. (1985). Epistemic dependence. *Journal of Philosophy, 82,* 335-349.

Hendrix, G.G. (1979). Encoding knowledge in partitioned networks. In N.V. Findler (Ed.), *Associative networks.* New York: Academic.

Hintikka, J. (1962). *Knowledge and belief.* Ithaca, NY: Cornell University Press.

Hintikka, J. (1967). Indexicals, possible worlds, and epistemic logic. *Noûs, 1,* 33-62.

Joshi, A., Webber, B., & Weischedel, R.M. (1984). Preventing false inferences. *Proceedings of the 10th International Conference on Computational Linguistics (COLING-84)* (pp. 134-138). Morristown, NJ: Association for Computational Linguistics.

Kant, I. (1929). *Critique of pure reason* (2nd ed.). (N. Kemp Smith, Trans.). New York: St. Martin's Press. (Original work published 1787)

Konolige, K. (1985). Belief and incompleteness. In J.R. Hobbs & R.C. Moore (Eds.), *Formal theories of the commonsense world.* Norwood, NJ: Ablex.

Lewis, D. (1979). Attitudes *de dicto* and *de se. Philosophical Review, 88,* 513-543.

Maida, A.S. (1985). *Selecting a humanly understandable knowledge representation for reasoning about knowledge.* International Journal of Man-Machine Studies, *22,* 151-161.

Maida, A.S., & Shapiro, S.C. (1982). Intensional concepts in propositional semantic networks. *Cognitive Science, 6,* 291-330.

Marburger, H., Morik, K., & Nebeli, B. (1984, June). The dialog system HAM-ANS: Natural language access to diverse application systems (Lecture presented at CSLI, Stanford University). Abstract in *CSLI Newsletter, 38,* 1-2.

Martins, J. (1983). *Reasoning in multiple belief spaces* (Tech. Rep. No. 203). Buffalo, NY: State University of New York at Buffalo, Department of Computer Science.

McCalla, G., & Cercone, N. (1983). Guest Editors' introduction: Approaches to knowledge representation. *Computer, 16 (October)* No. 10, 12-18.

McCarthy, J. (1979). First order theories of individual concepts and propositions. In J.E. Hayes, D. Michie, & L. Mikulich (Eds.), *Machine intelligence 9.* Chichester, England: Ellis Horwood.

Meinong, A. (1971). Über Gegenstandstheorie. In R. Haller (Ed.), *Alexius Meinong Gesamtausgabe* (Vol. II). Graz, Austria: Akademische Druck- u. Verlagsanstalt. (Original work published in 1904)

Minsky, M.L. (1968). Matter, mind, and models. In M. Minsky (Ed.), *Semantic information processing.* Cambridge, MA: MIT Press.

Montague, R. (1974). *Formal philosophy.* New Haven, CT: Yale University Press.

Moore, R.C. (1977). Reasoning about knowledge and action. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence (IJCAI-77)* (pp. 223-227). Los Altos, CA: Morgan Kaufmann.

Moore, R.C. (1980). *Reasoning about knowledge and action,* (Tech. Note No. 191). Menlo Park, CA: SRI International Artificial Intelligence Center.

Nilsson, N.J. (1983, Winter). Artificial intelligence prepares for 2001. *AI Magazine 4.4,* pp. 7-14.

Parsons, T. (1980). *Nonexistent objects.* New Haven, CT: Yale University Press.

Perry, J. (1979). The problem of the essential indexical. *Noûs, 13,* 3-21.

Perry, J. (1983). Castañeda on He and I. In J.E. Tomberlin (Ed.), *Agent, language, and the structure of the world.* Indianapolis, IN: Hackett.

Perry, J. (1985). Perception, action, and the structure of believing. In R. Grandy & R. Warner (Eds.), *Philosophical Grounds of Rationality* (Oxford: Oxford University Press).

Rapaport, W.J. (1976a). On *cogito* propositions. *Philosophical Studies, 29,* 63-68.

Rapaport, W.J. (1976b). *Intentionality and the structure of existence.* Unpublished doctoral dissertation, Indiana University.

Rapaport, W.J. (1978). Meinongian theories and a Russellian paradox. *Noûs, 12,* 153-180; errata, *Nous, 13*(1979), 125.

Rapaport, W.J. (1979). An adverbial Meinongian theory. *Analysis, 39,* 75-81.

Rapaport, W.J. (1981). How to make the world fit our language: An essay in Meinongian semantics. *Grazer Philosophische Studien, 14,* 1-21.

Rapaport, W.J. (1982). Meinong, defective objects, and (psycho-)logical paradox. *Grazer Philosophische Studien, 18,* 17-39.

Rapaport, W.J. (1984a). Critical notice of Routley 1979. *Philosophy and Phenomenological Research, 44,* 539-552.

Rapaport, W.J. (1984b). Belief representation and quasi-indicators (Tech. Rep. No. 215). Buffalo, NY: State University of New York at Buffalo, Department of Computer Science.

Rapaport, W.J. (1985a). To be and not to be. *Noûs, 19,* 255-271.

Rapaport, W.J. (1985b). Meinongian semantics for propositional semantic networks. *Proceedings of the Association for Computational Linguistics, 23,* 43-48.

Rapaport, W.J. (in press). Belief systems. In S.C. Shapiro (Ed.), *Encyclopedia of artificial intelligence.* New York: Wiley.

Rapaport, W.J., & Shapiro, S.C. (1984). Quasi-indexical reference in propositional semantic networks. *Proceedings of the 10th International Conference on Computational Linguistics (COLING-84)* (pp. 65-70).

Rich, E. (1979). *User modeling via stereotypes. Cognitive Science, 3,* 329-354.

Routley, R. (1979). *Exploring Meinong's jungle and beyond* (Canberra: Australian National University, Research School of Social Sciences, Department of Philosophy.

Schank, R.C., & Riesbeck, C.K. (Eds.). (1981). *Inside computer understanding.* Hillsdale, NJ: Erlbaum.

Shapiro, S.C. (1979). The SNePS semantic network processing system. In N.V. Findler (Ed.), *Associative networks.* New York: Academic.

Shapiro, S.C. (1981, July). *What do semantic network nodes represent?* Paper presented at Conference on Foundational Threads in Natural Language Processing, SUNY Stony Brook (SNePS Research Group Technical Note No. 7). Buffalo, NY: SUNY at Buffalo, Dept. of Computer Science.

Shapiro, S.C. (1982). Generalized augmented transition network grammars for generation from semantic networks. *American Journal of Computational Linguistics, 8,* 12-25.

Shapiro, S.C., & Rapaport, W.J. (in press). SNePS considered as a fully intensional propositional semantic network. In N. Cercone & G. McCalla (Eds.), *Knowledge representation.* Berlin: Springer-Verlag. Cf. *Proc. AAAI-86* (Los Altos, CA: Morgan Kaufmann), Vol. 1, pp. 278-283.

Stalnaker, R.C. (1981). Indexical belief. *Synthese, 49,* 129-151.

Wahlster, W. (1984, July). *User models in dialog systems.* Lecture presented at the 10th International Conference on Computational Linguistics (COLING-84), Stanford University.

Wiebe, J.M., & Rapaport, W.J. (1986). Representing *de re* and *de dicto* belief reports in discourse and narrative. *Proceedings of Institute of Electronic and Electrical Engineers.*

Wilensky, R., Arens, Y., & Chin, D. (1984). Talking to UNIX in English: An overview of UC. Communications of the Association for Computing Machinery, 27, 574–593.

Wilks, Y., & Bien, J. (1983). Beliefs, points of view, and multiple environments. *Cognitive Science, 7,* 95–119.

Woods, W.A. (1975). What's in a link: Foundations for semantic networks. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding.* New York: Academic.