MODELS AND MINDS: Knowledge Representation for Natural-Language Competence

Stuart C. Shapiro and William J. Rapaport

Department of Computer Science
and Center for Cognitive Science
State University of New York at Buffalo
Buffalo, NY 14260

shapiro@cs.buffalo.edu, rapaport@cs.buffalo.edu

Abstract

Cognitive agents, whether human or computer, that engage in natural-language discourse and that have beliefs about the beliefs of other cognitive agents must be able to represent objects the way they believe them to be and the way they believe others believe them to be. They must be able to represent other cognitive agents both as objects of beliefs and as agents of beliefs. They must be able to represent their own beliefs, and they must be able to represent beliefs as objects of beliefs. These requirements raise questions about the number of tokens of the belief-representation language needed to represent believers and propositions in their normal roles and in their roles as objects of beliefs. In this paper, we explicate the relations among nodes, mental tokens, concepts, actual objects, concepts in the belief spaces of an agent and the agent's model of other agents, concepts of other cognitive agents, and propositions. We extend, deepen, and clarify our theory of intensional knowledge representation for natural-language processing, as presented in previous papers and in light of objections raised by others. The essential claim is that tokens in a knowledge-representation system represent only intensions and not extensions. We are pursuing this investigation by building CASSIE, a computer model of a cognitive agent and, to the extent she works, a cognitive agent herself. CASSIE's mind is implemented in the SNePS knowledge-representation and reasoning system.

This paper is forthcoming in R. Cummins & J. Pollock (eds.), *Philosophical AI: Computational Approaches to Reasoning* (Cambridge, MA: MIT Press).

1 INTRODUCTION.

Knowledge representation is concerned, among other things, with the representation of information in artificial-intelligence (AI) computer systems. Few, if any, researchers claim at the outset to be working on an AI system whose abilities are co-extensive with human intellectual abilities, so most AI systems are designed to work on a narrower (if still broad) domain, such as vision, problem solving, robotics, or natural-language processing. It may be argued that in order to reach human-level abilities in any one of these areas, it is necessary to have human-level abilities in all the others. If that is true, then choosing an area of AI to work on is just a research strategy for working on general AI. In any case, the choice determines what one thinks are the problems to be worked on (initially) and the information that needs to be represented.

In this paper, we shall consider the information that needs to be represented by an AI system whose domain is general natural-language processing. By "natural-language processing", we mean the use of natural human languages (e.g., English) for the kind of communication with agents (human or computer) that humans engage in when they communicate with each other in those languages. In particular, we exclude the kinds of text processing that treat natural-language texts as uninterpreted strings of characters (such as producing a concordance), and we include both natural-language *generation* and natural-language *understanding*, for which we suggest the term *natural-language competence* (NLC). The quintessential NLC task is interactive dialogue, although other tasks (such as reading and summarizing a narrative or translating a text from one natural language to another) have been recognized by AI researchers as requiring the same abilities.

In order to focus on the information that an NLC system needs, it is useful to consider briefly, by way of contrast, the information needed by some other intelligent systems (remembering, however, that, as an NLC system is extended into a general intelligent system, it, too, will need such information). Robot systems (both for locomotion and for manipulation) and vision systems must operate in the here and now of the real world. The way the world really is (the subject of physics) is important to them, because they must operate in it. They must recognize multiple interactions with the same physical object as *being* interactions with the same physical object. This is the essence of the stereo vision problem: a mobile robot can go between two objects but must go around one; a manipulator needs at least two points of contact to grasp an object. On the other hand, although face-to-face dialogue often includes references to objects in the current environment, the essence of natural language is the ability to discuss the not-here and the not-now.

Distinguishing "knowledge representation" from "data storage" is not easy, but Brian Cantwell Smith's Knowledge-Representation Hypothesis is fairly well-accepted:

> Any mechanically embodied intelligent process will be comprised of structural ingredients that a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behaviour that manifests that knowledge. (Smith 1982: 33.)

As is often the case in AI, this takes key terms (such as "knowledge") on their pre-theoretic meaning, to be attributed (or not) to AI systems by the same criteria by which they are attributed to humans.

The declarative/procedural controversy (Winograd 1975) has fairly well died out, and it would seem that the "structural ingredients" in Smith's Knowledge-Representation Hypothesis, which are to be taken as the represented knowledge, could be either declarative or procedural. It is clear that the declarative knowledge of a system could be considered procedural to the extent that it plays a "causal and essential role in engendering . . . behaviour". However, the distinction is still useful if worded in a slightly different way than it has been in the past. Let us consider as *procedural knowledge* that knowledge that *only* plays such a causal role, and let us consider as *declarative knowledge* that knowledge that the system can discuss with another. So, for example, humans act as though they know the grammar of their native language (competence), but most cannot describe that grammar at all, and no one has yet been able to describe it completely. So we would say that people have procedural knowledge of the grammar of their native language, but not declarative knowledge of it. In this paper, we shall concentrate on discussing the declarative knowledge needed by a system for human-level NLC abilities, rather than the procedural knowledge needed for NLC. (For a discussion of the latter, see Weischedel 1986.)

2 INTENSIONAL KNOWLEDGE REPRESENTATION.

Insofar as an AI/NLC system is considered a model of a cognitive agent, the information represented in its "mind" consists of the beliefs, knowledge, and other intentional (i.e., psychological) attitudes of the agent, together with the objects, properties, and relations that those attitudes are directed to. (On the information represented, cf. Brachman & Smith 1980; on the notion of "mind", cf. Rapaport 1988a.)

Thus, cognitive agents, whether human or computer, that engage in natural-language discourse and that have beliefs about the beliefs of other cognitive agents must be able to represent objects the way they believe them to be and the way they believe others believe them to be. They must be able to represent other cognitive agents both as objects of beliefs and as agents of beliefs. They must be able to represent their own beliefs, and they must be able to represent beliefs as objects of beliefs.

A fair number of AI researchers have described systems, both theoretical and implemented, that address these issues. One remaining controversy is the number of levels of representation required to solve the problem. Consider a cognitive agent, C, with beliefs about another cognitive agent, O. Let "A" be the term C uses in its belief structure to refer to O. That is, if C believes that some property holds of O, then the term "A" actually occurs in the structure representing that belief. Here are some questions to consider:

- Does "A" refer to O extensionally, as an object-in-the-world, or only intensionally, as C's conceptualization of O?
- Are two terms, "A" and "A*", needed to represent the above two cases?
- Now suppose that "P" is the structure representing one of C's beliefs, Q, and that C believes that O also holds Q. Does it make sense for "P" and "A" to occur in C's structure representing the belief that O holds Q, without being quoted, primed, or otherwise modified?
- Suppose that C has another belief about the belief Q; how is "P" used to represent that belief?

In a series of earlier papers, we have argued that such a system must represent and reason about *intensional*, as opposed to *extensional*, entities (Maida & Shapiro 1982; Rapaport & Shapiro 1984; Rapaport 1985, 1986a; Wiebe & Rapaport 1986; Shapiro & Rapaport 1987a). Briefly, this means that, instead of storing information about *objects in the world*, as vision and robot systems must, NLC systems store information about *mental objects* that may or may not even exist.

Our theory of intensional knowledge representation answers the above questions as follows:

- "A" refers to O intensionally, as C's conceptualization of O.
- No additional term is needed to refer to O as an object-in-the-world.
- It does make sense for "P" and "A" to occur in the structure C uses to represent its belief that O holds Q, without modification.
- Similarly, "P" may be used without modification in the representation of other of C's beliefs about Q.

In Shapiro & Rapaport 1987a, we discussed the use of the SNePS propositional semantic-network processing system (Shapiro 1979) to build an NLC system according to our theory of intensional knowledge representation. This system, referred to as CASSIE (the *C*ognitive *A*gent of the *S*NePS *S*ystem—an *I*ntelligent *E*ntity) is a particular application of SNePS, with a particular set of arc labels and case frames, together with a generalized ATN parsing/generating grammar (Shapiro 1982) for a fragment of English. We refer to SNePS with these arcs, case frames, and grammar as SNePS/CASSIE.

In a series of papers, John Barnden has critiqued this theory. In Barnden 1986a, he claimed that, under it, there is no consistent way of dealing with sentence sequences such as:

(1) That John is taller than Mary is Kevin's favorite proposition.

(2) Bill believes Kevin's favorite proposition.

Here, Bill is O, "John is taller than Mary" is Q, (2) expresses the belief that O believes Q, and (1) expresses another belief about Q. He has also objected that a knowledge-representation system such as ours "imputes" "arcane" theories to cognitive agents, cannot quantify over intensions, and cannot represent the differences between an entity and an idea of that entity (Barnden 1986b, 1989).

In this essay, we extend, deepen, and, we hope, clarify, our theory of intensional knowledge representation implemented in CASSIE, and we discuss how CASSIE handles Barnden's challenges. We hope, thereby, to shed light on general issues of the modeling of cognitive agents sophisticated enough to have beliefs about the beliefs of other cognitive agents.

3 KNOWLEDGE-REPRESENTATION FORMALISMS.

3.1 Syntax and Semantics.

A knowledge-representation formalism, like any language, has a syntax and a semantics. The syntax specifies those parts of the language (such as "terms" and "sentences") that can be given meanings (interpretations) by the semantics. Larger syntactic parts (such as sentences) are formed from smaller syntactic parts by putting the smaller parts together in various structured ways with the use of punctuation. We presume that the reader is already familiar with this (a typical example for a first-order language is discussed in Rapaport 1987).

Several different knowledge-representation formalisms have been suggested for NLC systems. In the following sections, we survey a few of the major varieties, discussing their syntactic constructs and their intended domains of interpretation. (Other recent reviews of knowledge representation are Barr & Davidson 1981, Mylopoulos & Levesque 1984, Levesque 1986, and Kramer & Mylopoulos 1987. Recent collections of papers on knowledge representation relevant to NLC include Brachman & Levesque 1985, King & Rosner 1986, and Cercone & McCalla 1987.)

3.1.1 Conceptual Dependency Theory.

Conceptual Dependency theory (Schank & Rieger 1974, Schank 1975, Schank & Riesbeck 1981; cf. Hardt 1987) uses a knowledge-representation formalism consisting of sentences, called "conceptualizations", which assert the occurrence of events or states, and six types of terms: PPs—"real-world objects", ACTs—"real-world actions", PAs—"attributes of objects", AAs—"attributes of actions", Ts—"times", and LOCs—"locations". (The glosses of these types of terms are quoted from Schank & Rieger 1974: 378-379.) The set of ACTs is closed and consists of the well-known primitive ACTs PTRANS (transfer of physical location), ATRANS (transfer of an abstract relationship), etc. The syntax of an event conceptualization is a structure with six slots (or arguments), some of which are optional: actor, action, object, source, destination, and instrument. A stative conceptualization is a structure with an object, a state, and a value. Only certain types of terms can fill certain slots. For example, only a PP can be an actor, and only an ACT can be an action. Interestingly, conceptualizations themselves can be terms, although they are not one of the six official terms. For example, only a conceptualization can fill the instrument slot, and a conceptualization can fill the object slot if MLOC (mental location) fills the act slot. A "causation" is another kind of conceptualization, consisting only of two slots, one containing a causing conceptualization and the other containing a caused conceptualization.[1] Although, from the glosses of PP and ACT, it would seem that the intended domain of interpretation is the real world, the domain also must contain theoretically postulated objects such as: the "conscious processor" of people, in which conceptualizations are located; conditional events; and even negated events, which haven't happened.

3.1.2 The KL-ONE Family.

KL-ONE (Brachman & Schmolze 1985) and its descendents, KL-TWO (Vilain 1985), KRYPTON (Brachman, Fikes, & Levesque 1983), and Loom (MacGregor & Bates 1987) separate their formalisms into two sub-languages—the definitional (or terminological) component, called the "TBox", and the assertional component, called the "ABox" (KL-ONE's separation is more rudimentary than that of the later systems). The ABox consists of sentences in a restricted first-order logic (e.g., KL-TWO doesn't allow any quantifiers) and are taken to assert truths in the domain of interest. Terms in the ABox have existential import. That is, constants represent individuals that exist in the domain, and an existentially quantified sentence of the form $\exists x P(x)$ asserts that an individual satisfying P exists in the domain. Unary predicate symbols, relational predicate symbols, and function symbols in ABox sentences are themselves terms in the TBox. There are two types of terms in the TBox—concepts and relations (or roles). TBox sentences make assertions like, "A **family** is: a kind of **social-structure** with exactly 1 *male-parent* and this *male-parent* is a **man**; and

[1] Schank's notion of "causation", it should be noted, is such that two conceptualizations can "cause" *each other*. E.g., buying and selling events (both of which are ATRANSes) cause each other (cf. Schank & Riesbeck 1981: 17).

a kind of **social-structure** with exactly 1 *female-parent* and this *female-parent* is a **woman**; and a kind of **social-structure** all of whose *child*ren are **persons**." Here, **family**, **social-structure**, **man**, **woman**, and **person** are concept terms; *male-parent*, *female-parent*, and *child* are relation terms; all other parts of the sentence are punctuation; and the sentence itself is taken as the definition of **family**. This last is most important—the sentence cannot be either true or false; it is simply a definition. If you think that this definition is incorrect, you cannot say, "No, you're wrong; a family isn't like that." You can only say, "Well, your concept of **family** is different from mine." Also, defining **family** in the TBox doesn't say that there are any in the domain of interest. That can only be said in the ABox by an existential statement such as $\exists x \mathbf{family}(x)$ or by an ABox assertion such as **family**(Shapiro). TBox sentences about relations are also definitional. For example, we could use the relation *child* to define *son* as a *child* that is a **man**. NIKL (Moser 1983) is another descendent of KL-ONE and consists only of a TBox (in fact, it is the TBox of KL-TWO), so all sentences in NIKL are definitional.

3.1.3 SNePS.

The knowledge-representation formalism we are concerned with in this paper is SNePS/CASSIE. In this section, we give an introductory description of SNePS/CASSIE at the same level of detail we used for Conceptual Dependency and the KL-ONE family in the previous two sections.

The SNePS formalism consists of two parts, nodes and labeled, directed arcs. Nodes constitute the terms of the language; labeled, directed arcs are punctuation. Nodes are partitioned into atomic nodes, which have no arcs emanating from them, and molecular nodes, which do have arcs emanating from them. Every node represents some entity in the domain of discourse, and no two nodes represent the same entity. (We have called this the Uniqueness Principle.) The entity represented by a node is determined in one of two ways: it is determined *assertionally* by the sub-network connected to the node via arcs pointing into it; it is determined *structurally* by the sub-network connected to the node via arcs pointing out of it. (There are no cycles of arcs.) Once a node is created, it can get new arcs pointing into it, but the set of arcs pointing out of it can never change, so the structurally determined nature of a node is more characteristic of it than its assertionally determined nature. Notice that the semantics is not compositional, since the interpretation of an atomic node is not fixed independently of the nodes that get their structural interpretation by dominating it (having arcs that point out of them into it). In a sense, the semantics is circular—two nodes connected to each other each partially determine the other's meaning. A major escape from this circularity is the set of *sensory* nodes, which are atomic nodes representing lexical entries of the language that CASSIE uses to communicate with other cognitive agents (e.g., people) (cf. Rapaport 1988a). The entity represented by a molecular node is structurally determined semi-compositionally by the set of nodes it dominates[2] and by the labels of the arcs that connect it to them. For example, a node with a LEX arc to a sensory node represents the entity referred to by the lexeme that the sensory node represents, and a node with an OBJECT arc to a node i and a PROPERTY arc to a node j represents the proposition that the entity represented by i has the property represented by j. This is the sense in which the arcs are punctuation.

Notice that some nodes (both molecular and atomic) represent propositions (we will call these *propositional nodes*), but all nodes are terms of the SNePS language. Thus, molecular nodes, whether they represent propositions or what are usually considered to be individuals, correspond more to functional terms in other languages than to sentences. In this view, there is only one kind of sentence in SNePS. Furthermore, every propositional node has an *assertion tag* that can be on or off. If propositional node p represents proposition $[[p]]$, then p's assertion tag's being on represents the assertion, for CASSIE, that "I believe $[[p]]$". The assertion tag's being off represents the absence of such belief. Finally, some propositional nodes can be interpreted by the SNePS Inference Package (SNIP) as *rule nodes*. Rule nodes represent deduction rules and are used for node-based deductive inference (Shapiro 1978; Shapiro & McKay 1980; McKay & Shapiro 1981; Shapiro, Martins, & McKay 1982).

The ideas outlined in this section will be expanded upon in the rest of this paper.

[2] Only *semi*-compositionally, because it partially assertionally determines the meaning of them.

4 SEMANTICS.

4.1 Nodes, Mental Tokens, Concepts, and Actual Objects.

Above, we said that, instead of storing information about objects in the world, as vision and robot systems must, NLC systems store information about mental objects that may or may not even exist. CASSIE, as a cognitive agent implemented in SNePS, is an "interpreted automatic formal system" (to use Haugeland's (1985: 106) phrase). SNePS nodes are the tokens of CASSIE's mental language (cf. Fodor 1975, Stich 1983). As symbols of mental syntax, what do they denote? Our answer is that they denote elements of the domain that has been called "Aussersein": the domain of the objects of thought—the things we can think about, have beliefs about, etc., whether or not they exist or are true (Meinong 1904; Rapaport 1978, 1985; Shapiro & Rapaport 1987a). These elements can be called "mental objects" or "intensions". They are what we refer to as "concepts". So when we say that a node of SNePS/CASSIE represents a concept, we mean it in this way. Moreover, all representations of concepts are in CASSIE's mind. Thus, when CASSIE represents the beliefs of someone else, what is represented is CASSIE's representation of the beliefs of the other person, not the other person's actual beliefs. (We return to this in Section 5.2, below.)

What is the relation between a concept (an intensional object in Aussersein) and objects in the world? Some concepts have objects in the world that *correspond to* them (Rapaport 1978 refers to the objects as the "Sein-correlates" of the concepts); some have one; some have more than one. Some have none: for instance, concepts of unicorns and square circles have no objects corresponding to them in the real world. The concept of the coffee cup that I[3] habitually use in and around my office has one object that corresponds to it. The concept of the straight-backed teak chair with black seat that I[4] have in my house has eight objects that correspond to it.

I[5] believe that I can recognize my coffee cup, but it may be that there is one exactly like it. If, some night, someone were to replace my coffee cup with one that looked exactly like it, I might not notice the difference (cf. The Ballad of Shakey's Pizza Parlor in Dennett 1982: 53-60). In that case, I would take the new cup to be my cup, the object that corresponds to the concept denoted by the mental token I have been using for that concept for years, and I would not create a new mental token for it. On the other hand, I know I cannot distinguish my eight chairs at home, and have no beliefs about any individual one.

What do I mean by "my concept of X", and how does that relate to "your concept of X"? If my beliefs about X coincide exactly with your beliefs about X (no matter how long we discuss X, we find no disagreement), then our concepts of X are the same intensional objects (although there would be two nodes representing them, one in my mind, one in yours). However, if we disagree in the slightest about X, then our concepts of X are different intensional objects in Aussersein, even if one and the same object in the real world corresponds to them. This also applies to "CASSIE's concept of X" and "OSCAR's concept of X", where CASSIE and OSCAR (the *O*ther *S*NePS *C*ognitive *A*gent *R*epresentation) are two different SNePS cognitive agents.[6]

4.2 Epistemological Ontology.

What sorts of things are there in Aussersein? Since this is a question of what there is among mental objects, it is a question of epistemological ontology (cf. Rapaport 1985/1986). It can be answered by considering the naive ontology of things talked about in natural language—what the world would be like if it fit our language (Rapaport 1981; cf. Hobbs 1985).

One sort of mental object is people. I have beliefs about George Bush; so I have a concept of George

[3]When "I" is used, the reader is invited to use a new mental token for a concept that is co-extensional with one of the two authors, but the reader does not know which.

[4]The reader may identify the referent of this "I" with the previous one, or may create a new mental token, again co-extensional with one of the two authors. In the latter case, the two "I's" might or might not refer to the same author, but in the former case, the reader is deciding that they refer to the same one, but still has not committed to which one.

[5]This now, surely, refers to the same author as one of the two previous "I's," and the reader should have no trouble deciding which.

[6]Our OSCAR is, perhaps, a distant cousin of John Pollock's (1988) Oscar; cf. Rapaport, Shapiro, & Wiebe 1986.

Bush. My concept of George Bush includes beliefs that he is currently President of the United States, was Vice President of the U.S., before that was Director of the CIA, etc. There is an object in the real world (at least, I believe so) that corresponds to the concept of a person named George Bush having these three properties. Depending on the rest of my beliefs about George Bush, there may or may not be an object in the real world corresponding to it (cf. Rapaport 1978).

Another sort of mental object is mental acts, such as acts of believing. Our concept of believing is being discussed in this essay, and there is a vast literature discussing other people's concepts of believing. Whether one believes that there is an object in the real world corresponding to believing depends on one's ontology (i.e., one's *non*-epistemological ontology), and the two authors of this paper sometimes disagree on that.

Another sort of mental object is propositions. Propositions are sometimes called "beliefs", "facts", or "truths", depending on what other beliefs the speaker has about the proposition being discussed. The ontological status of objects corresponding to propositions is as controversial as that of objects corresponding to beliefs.

There are other sorts of mental objects. Some are: properties, sentences, numbers, numerals, and truth values. These are all different sorts in the sense that we have different kinds of beliefs about them and believe that different kinds of properties are relevant to them.

When we discuss people, we refer to them by using a proper name, a definite noun phrase, or, sometimes, an indefinite noun phrase. We can refer to Ronald Reagan as "Ronald Reagan" or "Nancy Reagan's favorite actor". A particular understander at a particular time will take these referring expressions to refer to his or her own concept of Ronald Reagan, and may react accordingly.

When we discuss propositions, we often use the phraseology, "that ⟨clause⟩". For example, in the U.S. Declaration of Independence, we find, "We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness", and in Lincoln's Gettysburg Address we find, "a new nation conceived in liberty and dedicated to the proposition that all men are created equal." We could also refer to a proposition using a definite noun phrase, e.g., "the first self-evident truth mentioned in the Declaration of Independence" or "Lincoln's favorite proposition". Whether the referent of "the first self-evident truth mentioned in the Declaration of Independence" *is the same as* the referent of "Lincoln's favorite proposition" depends on the particular beliefs of a particular understander.

People, believings, and propositions (those people, believings, and propositions that we have beliefs about) are denizens of the same universe—Aussersein. If CASSIE believes that Lincoln believed that all men are created equal, part of CASSIE's Lincoln concept is that he participates in a believing act (her concept of believing) directed to the proposition that all men are created equal (her concept of that proposition).

5 TOPICS IN KNOWLEDGE REPRESENTATION FOR NATURAL-LANGUAGE COMPETENCE.

In this section, we present short discussions of a series of topics we believe to be important for knowledge representation for NLC. We discuss these topics in terms of SNePS/CASSIE, not because we believe that they are only relevant to her, but because we feel that we can express ourselves most clearly using her as a model.

5.1 Knowledge and Belief.

CASSIE needs to be able to represent her own beliefs. She should be able to represent them both *explicitly* (or consciously)—she must be able to *say* "I believe that P"[7]—and *implicitly* (or subconsciously)—she must be able to say "P" (and it must be the case that *we* can say, truthfully, that she believes that P). The latter is

[7]She ought to be able to say this even if, in fact, she *doesn't* believe that P—perhaps she only *thinks* (mistakenly) that she believes that P.

handled by *asserting* that P (cf. Sect. 3.2.3, above). The former is handled by having a node that represents CASSIE's self-concept (it has an "I"-pointer, similar to the "now"-pointer that marks (CASSIE's concept of) the current time (Rapaport, Shapiro, & Wiebe 1986; on the "now"-pointer, cf. Almeida & Shapiro 1983, Almeida 1987).

She must also be able to represent the beliefs of others: (a) *de re* belief reports ("John believes of Lucy that she is sweet", where 'Lucy' is the name that *CASSIE* uses to refer to the individual whom John believes to be sweet, whether or not that is how John would refer to her), (b) *de dicto* belief reports ("John believes that the girl next door is sweet", where 'the girl next door' is how CASSIE understands *John* to characterize her), and (c) quasi-indexical *de se* belief reports ("John believes that he* is rich", where the use of 'he' by CASSIE expresses what John would express with 'I'; cf. Castañeda 1966, Rapaport 1986a).[8] These are handled as shown in Figures 1–3. In a linear, predicate language, these are, roughly,

(*de re*) Name(b1, 'John') & Name(b2, 'Lucy') & Believes(b1, Sweet(b2))

(*de dicto*) Name(b1, 'John') & Believes(b1, The-girl-next-door(b2)) & Believes(b1, Sweet(b2))

(*de se*) Name(b1, 'John') & Believes(b1, Rich(b1))

Note that a sentence expressing a belief report can be *interpreted* as *de re* or *de dicto*, and the interpretation determines how the report will be represented. Once represented, the terms '*de re*' and '*de dicto*' no longer apply (cf. Wiebe & Rapaport 1986). (The representation of a belief report interpreted *de dicto* consists of two, linked representations of *de re* belief reports. For a critique of this, see Wyatt 1989. The representation of *de se* belief reports given here differs from that given in Rapaport & Shapiro 1984 and Rapaport 1986a. The reasons for the change are detailed in Rapaport, Shapiro, & Wiebe 1986, and are related to our representation for knowledge reports and the veridicality principle, described next.)

Arguably, there are also *de re*, *de dicto*, and *de se* knowledge reports; these are represented similarly, replacing the sensory-node labeled with the lexeme 'believe' by a sensory-node labeled 'know'. A rule of veridicality (that what one knows is true) allows CASSIE to infer that if she believes that John knows that P, then she believes that P. That rule is shown in Figure 4; roughly, it is:

(Veridicality)

(CASSIE believes that A knows of N that F) $\rightarrow$
(CASSIE believes that A believes of N that F) & (CASSIE believes that N is F)

(For a detailed justification of this rule, see Rapaport, Shapiro, & Wiebe 1986.)

5.2 Fully Intensional Knowledge Representation in SNePS.

In order to distinguish the sort of intensional knowledge representation that we advocate, it will be useful to compare our theory with those of Hector-Neri Castañeda (1972, 1975ab, 1977) and Gottlob Frege (1892). We choose these, since, like ours, they are "fully intensional" in the sense that they do not use any notion of possible worlds. (There are, of course, possible-worlds analyses of Fregean theories, but Frege himself did not use them.)

One standard way to see how a semantic theory works is to test it on various "puzzles" (cf. Russell 1905). We shall give CASSIE the following puzzle (from Castañeda 1989; cf. Castañeda 1984):

[8] The '*' marks a quasi-indexical, or "logophoric" pronoun; cf. Sells 1987, Zubin in Rapaport, Segal, et al. 1989.

CASTAÑEDA'S PUZZLE

The following seven statements, according to standard principles of logic, imply a contradiction:

(1) At the time of the pestilence, Oedipus believed that: Oedipus's father was the same as his own father but the previous King of Thebes was not the same as his own father.

(2) Oedipus's father was the same as the previous King of Thebes.

(3) It was not the case that at the time of the pestilence Oedipus believed that: the previous King of Thebes was the same as his own father but the previous King of Thebes was not the same as his own father.

(T1) For any individuals x and y: if x is (genuinely or strictly) identical with y, then whatever is true of x is true of y, and vice versa.

(T2) The sentential matrix occurring in (1) and (3), namely: 'at the time of the pestilence, Oedipus believed that: ___ was the same as his own father but the previous King of Thebes was not the same as his own father', expresses something true of (a property of) the individual denoted by the singular term that by filling the blank in the matrix produces a sentence expressing a truth.

(T3) The expression 'was the same as' in (2) expresses genuine or strict identity.

(T4) The singular terms 'the previous King of Thebes' and 'Oedipus's father' have exactly the same meaning and denotation in both direct and indirect speech.[9]

Frege's theory can resolve the contradiction by accepting (T1)-(T3) and rejecting (T4). The rejection of (T4) is accomplished by his introduction of the sense-reference distinction: On his theory, the singular terms mentioned in (T4) differ in sense and reference in direct and indirect speech.

Castañeda's theory resolves the contradiction by accepting (T1), (T2), and (T4), but rejecting (T3). The meaning of a term is always a "guise" (a kind of intensional entity that is "located" in the actual world); thus (T4) is maintained. But 'is the same as' is taken as ambiguous among several different kinds of "sameness" relations, chief of which are "consubstantiation" (*very* roughly, co-extensionality) and "consociation" (*very* roughly, co-extensionality within an intensional context). In particular, (2) is taken as expressing the consubstantiation of two guises. (For details, see Castañeda 1972, 1975b, 1977, and, especially, 1975a; cf. Rapaport 1978, 1985.)

On our SNePS/CASSIE theory, however, all of (T1)-(T4) can be accepted without contradiction. CASSIE interacts with other cognitive agents in (fragments of) natural language, interpreting each sentence in light of her previous beliefs (cf. Shapiro & Rapaport 1987a, Rapaport 1988a). In this regard, our theory is similar to Hans Kamp's Discourse Representation Theory (Kamp 1984 and forthcoming; Asher 1986, 1987). Imagine CASSIE being told sentences (1)-(3), in that order. There are several ways she can interpret the descriptions in these sentences depending on her previous beliefs. Figure 5 shows one version of a SNePS semantic network of concepts and propositions representing CASSIE's beliefs after hearing (3).

After understanding (3), CASSIE believes the propositions represented by the following molecular nodes:

[9](T4) is ambiguous. The intended reading is that 'the previous King of Thebes' has exactly the same meaning and denotation in both direct and indirect speech, and so does 'Oedipus's father', not that the two descriptions have the same meaning and denotation as each other.

- m2 (that someone—namely, the individual represented by atomic node b1—is named 'Oedipus')
- m5 (that Oedipus believes that someone—namely, the individual represented by node b2—is named 'Oedipus')
- m8 (that Oedipus believes that someone—namely, the individual represented by node b3—is the father of the individual he believes to be Oedipus, i.e., of the individual represented by b2)
- m9 (that the individual represented by node b4 is Oedipus's father)
- m10 (that Oedipus believes that the individual represented by b4 is his father)—i.e., CASSIE believes that Oedipus believes that Oedipus's father is the same as his own father
- m16 (that something—namely, the individual represented by node b6—is named 'Thebes')
- m17 (that Oedipus believes that b6 is named 'Thebes')
- m18 (that something, b5, is Thebes's previous king)
- m19 (that Oedipus believes that b5 is Thebes's previous king)
- m20 (that b4 = b5, i.e., that Oedipus's father is Thebes's previous king)
- m23 (that Oedipus believes that m22, i.e., that Oedipus believes *both* that m11 (i.e., that b3 = b4, i.e., that Oedipus's father is his own father) *and* that m21 (i.e., that it is not the case that m20, i.e., that b4 $\neq$ b5, i.e., that his own father $\neq$ Thebes's previous king)—i.e., CASSIE believes that Oedipus believes that Thebes's previous king is not the same as his own father

(We are ignoring temporal information for simplicity, but without loss of generality; the representation of temporal information in SNePS is discussed in Almeida & Shapiro 1983, Almeida 1987.)

Note that there is no contradiction in this figure. To see why, consider each of (T1)-(T4):

(T1) In SNePS, there are two ways to represent "sameness": (a) by following the Uniqueness Principle (cf. Sect. 3.2.3, above) and using a single node, or (b) by using two nodes and asserting an "EQUIV"-proposition about them (such as propositional nodes m11 and m20 of Figure 5; cf. Maida & Shapiro 1982, Rapaport 1986a). Only the former is genuine or strict identity; the latter is co-extensionality within a belief space (corresponding, roughly, to Castañeda's "consociation"). In SNePS, A is "true of" B if A is asserted of B. But if x and y are genuinely or strictly identical (i.e., are the same node), then surely whatever is true of one is true of the other. So SNePS/CASSIE accepts (T1).

(T2) The hidden agenda behind (T2) concerns referential opacity and transparency. Thesis (T2) asserts that the blank in the quoted context is in a referentially transparent position. Now, SNePS/CASSIE distinguishes between *de re* and *de dicto* belief reports (cf. Sect. 5.1, above). So there are really two versions of (T2) for us to consider. On the *de re* version, the issue concerns the network fragment that consists of (a) the representation of a *de re* report of the appropriate form minus (b) the node representing whatever fills the blank: Is this fragment asserted of whatever is represented by that node? In SNePS/CASSIE, it is; so SNePS/CASSIE accepts the *de re* version of (T2). On the *de dicto* version, the issue concerns the network fragment that consists of (a) the representation of a *de dicto* report of the appropriate form minus (b) the node representing whatever fills the blank: Is this fragment asserted of that blank filler? But recall that we represent *de dicto* belief reports using two representations of *de re* belief reports (cf. Sect. 5.1, above). So, the answer here is the same as before: SNePS/CASSIE accepts the *de dicto* version of (T2), also. Figure 5 shows the *de dicto* interpretation of (1) with respect to "Oedipus's father," so the node representing the individual denoted by the term filling the blank is b3.

(T3) Since CASSIE hears (2) *after* hearing (1), *she already has separate concepts of Oedipus's father* (represented by node b4 of Figure 5) *and the previous King of Thebes* (represented by node b5). So her representation of (2) has to be a proposition asserting that those nodes are "EQUIV"alent (represented by node m20). Hence, in this case, SNePS/CASSIE rejects (T3). However, if CASSIE had heard (2) *before* (1), her interpretation of (2) *could* be (and there is psychological evidence that it *should* be: see Anderson 1977, 1978; cf. Maida & Shapiro 1982) *two* propositions asserting of a *single* entity both that *it* was Oedipus's father and that *it* was the previous King of Thebes. In that case, SNePS/CASSIE would accept (T3).

(T4) We now come to the heart of the matter (cf. Sect. 4.1, above):

> All nodes in the semantic network representing CASSIE's mind represent concepts in CASSIE's mind only. They never represent concepts in any other cognitive agent's mind. Concepts that are represented as being in another cognitive agent's belief space are just CASSIE's concepts about which she has beliefs that the other cognitive agent has beliefs about them.

(We have made this methodologically solipsistic point in Wiebe & Rapaport 1986. Wilks (1986: 267) makes a similar point, calling this "recursive cognitive solipsism".) Thus, (T4) is too simplistically stated to be either accepted or rejected. If the understander takes the believer to have the same concept of the referent of the singular term as the understander herself, she will use exactly the same node for this occurrence in indirect speech as she uses for occurrences in direct speech. (Figure 5 shows CASSIE making this interpretation for both "Thebes" and "the previous King of Thebes".) However, if the understander chooses to allow for the believer's having a different concept of the referent of the singular term, then she will use a different node, and this occurrence in indirect speech will not have exactly the same meaning as an occurrence of the term in direct speech. (Figure 5 shows CASSIE making this interpretation for the second occurrence of "Oedipus" and for the occurrence of "Oedipus's father" in (1).)

Thus, we do not, and do not need to, reject any of (T1)-(T4) (though we do reinterpret some of them).

5.3 Belief Spaces.

We define a believer's *belief space* to be the set of nodes (and, by extension, the concepts represented by them) dominated by propositional nodes representing propositions believed by the believer. All the nodes in Figure 5 are in CASSIE's belief space. All the nodes in Figure 5 dominated by (and including) nodes m4, m7, m9, m16, m18, and m22 are in the belief space of CASSIE's concept of Oedipus, referred to as (Oedipus CASSIE) in the notation of Rapaport & Shapiro 1984, Rapaport 1986a, and Wiebe & Rapaport 1986. We say "CASSIE's concept of Oedipus" rather than just "Oedipus", because Oedipus's belief space is in his own mind, while all nodes in Figure 5 are in CASSIE's mind. (Oedipus CASSIE)'s belief space contains the nodes representing concepts that CASSIE believes to be in Oedipus's belief space. That is the sense in which (Oedipus CASSIE)'s belief space is CASSIE's model of Oedipus's belief space.

All nodes in (Oedipus CASSIE)'s belief space are also in CASSIE's belief space. *Some* of the nodes in (Oedipus CASSIE)'s belief space are such that CASSIE and (Oedipus CASSIE) have nearly the same beliefs about them. That is the case when CASSIE believes that Oedipus has a mental token representing an intensional object in Aussersein that closely resembles the intensional object that the node represents for CASSIE. For example, both CASSIE and (Oedipus CASSIE) believe the proposition represented by node m16 in Figure 5. Thus, CASSIE believes that Oedipus has nearly the same concept of Thebes (node b6) as she does. Similarly, CASSIE believes that she and Oedipus closely agree on the concept of the previous King of Thebes (node b5), as well as the concept of the being-the-father-of relation (node m6), etc. The reason that this is close agreement rather than exact sharing of concepts is that, for example, CASSIE's concept of Thebes includes certain beliefs about its previous king that are not shared by (Oedipus CASSIE).

When CASSIE allows for the possibility that Oedipus and she might have very different concepts of an individual satisfying some description, she uses two nodes, one just in her belief space and one in

(Oedipus CASSIE)'s belief space. For example, in Figure 5, node b1 represents Oedipus for CASSIE, whereas b2 represents Oedipus for (Oedipus CASSIE). Only in this way can the sentence "Oedipus believed that Oedipus's father was the same as his own father" say anything non-tautologous. "His own father" refers to (Oedipus CASSIE)'s father, represented by b4, whereas this occurrence of "Oedipus's father" refers to (Oedipus Oedipus CASSIE)'s father, represented by b3. (Oedipus CASSIE)'s belief that they are the same is represented by node m11.

It might be noticed that Figure 5 does not show m11 as being believed by (Oedipus CASSIE). Instead, (Oedipus CASSIE) believes m22, which is the conjunction of m11 and m21. Reasoning within a belief space may be carried out by the SNePS Belief Revision system (SNeBR; Martins & Shapiro 1983, 1988), assuming that all believers accept SNeBR's rules of inference. A SNeBR belief space, consisting of a set of hypotheses and all propositions derived from them, may be initialized with all propositions believed by (Oedipus CASSIE) and all propositions believed by CASSIE to be "common knowledge". SNeBR's conclusions, which in this case would include m11, could then be installed into (Oedipus CASSIE)'s belief space by making them the objects of believings by (Oedipus CASSIE). This is another way in which CASSIE can use (Oedipus CASSIE)'s belief space as a model of Oedipus's belief space.

As another example, consider Figure 6, which shows CASSIE's belief space after she has processed the two sentences of Barnden's challenge (cf. Sect. 2, above). As before, nodes with assertion tags (exclamation marks) represent beliefs held by CASSIE. All the nodes in Figure 6 are in CASSIE's belief space. The nodes in Figure 6 dominated by (and including) node m6 are in the belief space of CASSIE's concept of Bill, i.e., (Bill CASSIE). We say "CASSIE's concept of Bill" rather than just "Bill", because Bill's belief space is in his own mind and all nodes in Figure 6 are in CASSIE's mind. (Bill CASSIE)'s belief space contains the nodes representing concepts that CASSIE believes to be in Bill's belief space. That is the sense in which (Bill CASSIE)'s belief space is CASSIE's model of Bill's belief space.

All nodes in (Bill CASSIE)'s belief space are also in CASSIE's belief space. However, CASSIE and (Bill CASSIE) may hold different beliefs about them. For example, since CASSIE gave a *de re* reading to "John is taller than Mary", she used nodes b1 and b2 to represent John and Mary, respectively. Node b1 represents the individual CASSIE believes to be named "John" and whom she believes that Bill believes to be taller than the individual represented by b2. CASSIE has no beliefs concerning the name that Bill uses when referring to the person represented by b1. Similarly, node m6 represents the proposition that CASSIE would express as "John is taller than Mary," and that she believes that Bill believes, although she, herself, doesn't believe it.

5.4 Concepts of Concepts: CASSIE as Cognitive Scientist.

Now that we have explicated our understanding of the relationships among mental tokens, nodes, concepts, actual objects, concepts of different cognitive agents, and concepts in different belief spaces, we will briefly discuss concepts of concepts.

The first thing to note is that we have, in fact, not yet discussed concepts of concepts. We have, however, discussed concepts. That is, this paper is intended to explicate our concept of concepts. CASSIE has heretofore had beliefs about people, properties, believings, propositions, etc., but she has not had any beliefs about concepts. If we started discussing cognitive science with CASSIE, we could give her such beliefs. She would then develop a concept of concepts, and might have a theory (a coherent set of beliefs) of how concepts relate to people. In this way, concepts are just another sort of mental object on the same level as intensional people, believings, and propositions. Whether concepts in Aussersein have objects in the real world corresponding to them is another controversy, which we do not want to pursue.

It is an intriguing idea that we could give CASSIE a theory of concepts different from the one we employed in implementing her. There would be no way for her to tell that her theory was not in accord with the facts, unless she made a prediction about her own behavior that was not confirmed, because she does not have access to the data structures used to implement SNePS.

If CASSIE thought about the issues raised in this paper, she might have nodes representing her concept of Oedipus, her concept of her concept of Oedipus, her concept of OSCAR, her concept of Oedipus as she believes OSCAR thinks of him, and her concept of OSCAR's concept of Oedipus. These would be different nodes, since they represent different intensional objects. (We return to this issue in our reply to Barnden in Sect. 6, below.)

5.5 Plans.

Plans are mental objects that need to be represented in NLC systems. Plans can be discussed, reasoned about, formulated, and followed.

It may be thought that any AI system that can represent and use rules, and that can evaluate a predicate by computation rather than by inference, can just use rules to represent plans. For example, Prolog is such a system, and one can view the Prolog rule,

$$p \ :- q, r, s.$$

either as the rule, "p is true if q, r, and s are true", or as the plan, "If you want to do p, first do q, then do r, and then do s". SNePS is also such a system, but in SNePS it is quite clear that rules can only be used as plans for reasoning, not as plans for acting.

To understand the problem, consider the Prolog rule,

$$p \ :- q, r, q.$$

This is bizarre as a rule—once q is determined as true, why bother to do so again? However, as a plan for acting, it is reasonable—there are many plans that require performing the same act more than once (for an appropriate meaning of "the same" applied to acts). The strict left-right depth-first control structure of most implementations of Prolog guarantees that the subgoal q is activated twice regardless of whether the rule is supposed to be a reasoning rule or a plan for acting. We can see that the use of rules as acting plans in Prolog depends on a pun: ',' is simultaneously interpreted as the logical operator *AND* and as the programming-language *sequence* operator.

SNePS does not make this pun. The corresponding SNePS rule is:

$$q \vee\!\!\!\rightarrow\ [r \vee\!\!\!\rightarrow\ [q \vee\!\!\!\rightarrow\ p]]].$$

(where '$\vee\!\!\!\rightarrow$' is "or-entailment"; cf. Shapiro 1979, Shapiro & Rapaport 1987a). Although this specifies the same order of subgoal triggering as the Prolog version, SNIP would not attempt to work on q the second time. It would just reuse the results of the first effort. Therefore, SNePS rules cannot be used as plans for acting.

Is this a deficit in SNePS? No, for two reasons. First, this refusal to work on the same subgoal more than once is not only efficient, it is one of the mechanisms that allow SNIP to work with unrestricted recursive rules (Shapiro & McKay 1980, McKay & Shapiro 1981), a facility Prolog does not have. Second, the Prolog pun amounts to a semantic confusion between the logical notion of truth (or belief) and the practical notion of action.

People can explain plans to each other using natural language. For example, a student might be told, "To sum some numbers, choose an accumulator variable, set it to zero, and then successively add the numbers to it." This sentence expresses a plan, using certain grammatical structures to express the structure of the plan. The student's understanding of the plan can then be tested by asking questions about it. For example, one might ask, "What should the accumulator variable be initialized to when summing a sequence of numbers?".

Of course, an excellent way of demonstrating an understanding of a plan is to follow the plan in appropriate circumstances. For an NLC system, the major action carried out is the generation of language, so one might expect an NLC system to be able to understand plans about how to express itself in natural language, reason about such plans, discuss them, and use them. The AI literature on planning is vast (see Vere 1987), but we know of no current NLC system that does all this.

5.6 Truth.

We have persistently avoided providing a standard Tarskian denotational semantics for our networks (although nothing prevents us from doing so, in the manner, e.g., that Kamp 1984 does for his theory). For one reason, this is because the entire network is definitional (like a NIKL network; cf. Sect. 3.2.2, above). For another, we are more interested in *CASSIE's representation* of the external world than in either the external world itself or the *correctness* of her representation of it. (This is another manifestation of our methodological solipsism.)

Consider Figure 5, again. Node m16, by its syntactic position in the network, represents a proposition that there is something named "Thebes". This something has the properties represented by the rest of the network connected to it, and subtly changes as the network grows (cf. the "pegs" of Landman 1986). One property is that the individual represented by b5 was its previous king. Again, this may be true, false, or just confused. It may be that the previous king was in some ways alike, and in some ways different from, CASSIE's concept of him. Would that make m18 false? Would it make m16 false?

The only construct in SNePS/CASSIE of any assertional import is the device that declares certain propositions (such as the one represented by m16) to be believed by CASSIE, and *these assertions are true by virtue of being made*—CASSIE is just built that way.

If CASSIE utters a sentence generated from a proposition she believes, we might agree or disagree with that sentence. Just as with a person, we might say the sentence is true or false, or we might say that CASSIE is confused: "Your concept of Thebes is somehow different from mine. Let's discuss it."

6 REPLIES TO BARNDEN'S OBJECTIONS.

John Barnden (1986ab, 1989) has made several points that he presents as clarifications and/or objections to our intensional knowledge-representation theory as expressed in Maida & Shapiro 1982 and Rapaport & Shapiro 1984 (however, the reader should consult Rapaport 1986a and Shapiro & Rapaport 1987a for more recent formulations). We feel that some of Barnden's points are valid complaints about lack of clarity (or timidity) in our earlier papers, some stem from his confusion about our theory, and some stem from actual differences in our theories. In this section, we summarize Barnden's points and reply to them.

Barnden wonders whether we want our nodes to "ambassadorially represent" intensions or objects in the world, or denote one or the other (1986a: 412). We are not sure we understand what Barnden means by "ambassadorially represents," but suspect that we would accept the idea that we intend nodes to ambassadorially represent intensions. We would not, however, accept Barnden's conclusion that, therefore, they *denote* objects in the world. Our ideas here were spelled out in Section 4.1, above.

Barnden wonders whether we want our network to model a cognitive agent, to be used by a cognitive agent, or to be a theoretical tool (1986a: 412). The answer to this is: all three. It is a theoretical tool for understanding cognition; it uses the methodology of AI, in which one of the best theoretical tools is a computer model; and, in that sense, it is also a model of the mind of a cognitive agent. However, sometimes, a model of X *is* an X, and we believe that this is one of those cases (see Section 7, below). Thus, the network also is used by a cognitive agent—CASSIE.

Barnden says that belief is a relation between an extensional person and an intensional proposition (1986a: 412-413) and that, therefore, "in determining what a proposition denoted by a node states, we sometimes 'dereference' the concepts denoted by argument nodes and sometimes we do not" (Barnden 1986a: 413). Instead, we believe that belief is an intentional stance ascribed to people (see Dennett 1978) and therefore is a relation between an intensional person and an intensional proposition (see Section 5.3, above).

Barnden goes on to suppose that we would have practical problems building "a system that translates network fragments into natural-language statements. ...[W]e do not want the language generator coming out with a statement to the effect that a *concept of* Bill believes something, or to the effect that Bill believes some *truth-value*" (Barnden 1986a: 413; italics in original). That our system has no such problems may be seen from the CASSIE conversations below and in Shapiro & Rapaport 1987a. See also the discussion of natural-language expressions that refer to intensional people and propositions in Section 5.3, above, and the discussion of concepts of concepts in Section 5.4, above.

As noted in Section 2, Barnden suggests that we have no consistent way of dealing with sentence sequences such as

(1) That John is taller than Mary is Kevin's favorite proposition.

(2) Bill believes Kevin's favorite proposition.

because the node representing the proposition would have to be dereferenced in one context and not in the other (1986a: 413-414).

However, we have given these sentences to CASSIE, and the conversation went as follows:

: That John is taller than Mary is Kevin's favorite proposition.
I understand that that John is taller than Mary is Kevin's favorite proposition

: Bill believes Kevin's favorite proposition.
I understand that Bill believes of John that he is taller than Mary.

Input sentences are on the lines beginning with the ":" prompt. The other lines are CASSIE's responses. CASSIE's response to a statement is the phrase "I understand that" prefixed to a sentence that she generates from the belief structure formed while analyzing the input statement. CASSIE's belief space after hearing these two sentences is shown in Figure 6. Node m6 is built when CASSIE understands the subject clause of sentence (1). Then it is given the property of being Kevin's favorite proposition. When CASSIE reads "Kevin's favorite proposition" in sentence (2), she takes it referentially to refer to the proposition represented by m6, and attaches it as the object of Bill's believing. Although CASSIE believes that the proposition represented by m6 is Kevin's favorite proposition, she does not express it that way, because she prefers to express propositions by giving their propositional content. The sentence comes out "...believes of John that he is taller than Mary" rather than "...believes that John is taller than Mary", because that is CASSIE's way of explicitly showing that a *de re* reading of "John" is appropriate.

A more general problem that Barnden sets for knowledge-representation systems is what he calls the "imputation problem" (1986b). Consider the following:

(S) Mike believes that the water is boiling.

(S′) Mike believes that the water is forcibly expelling some water vapor.

and suppose that boiling = forcibly expelling some water vapor. Clearly, one of (S), (S′) might be true while the other is false. Now suppose that X is an NLC program that represents 'The water is boiling' in its knowledge-representation language as:

(F) $\exists v[forcibly\text{-}expel(w, v) \wedge is\text{-}water\text{-}vapor(v)]$

According to Barnden, if X then represents (S) as:

(G) $B(mike, \exists v[forcibly\text{-}expel(w, v) \wedge is\text{-}water\text{-}vapor(v)])$

then X *imputes* its own theory of boiling to Mike. (One might say that (G) is a *de dicto* belief report; cf. Sect. 5.1, above.) Barnden sees this as a problem insofar as such an imputation is "harmful" (not all are), and he proposes a strategy for ameliorating the imputation problem in the case of representation schemes intended for ... [cognitive modeling]:

> STRATEGY 1. Make sure that all explications that the representation scheme induces are cognitively reasonable for human beings. In particular, do not explicate propositional attitudes in terms of arcane theoretical constructs. (Barnden 1986b: 341.)

According to Barnden (1986b: 345f), this is the strategy pursued by SNePS even though an imputation problem remains: Consider

(S15) Mike believes [*de re*] that Jim's mother is clever

(which we prefer to express as 'Mike believes of Jim's mother that she is clever'), and suppose

(*) George believes of Jim's mother that Mike believes that she is clever.

Now, how should these be represented by SNePS/CASSIE? The representation of (S15) might be as shown in Figure 7a or 7b. That is, roughly,

(S15a) Believes(Mike, Clever(b3)) & Mother-of(Jim, b3)

i.e., Mike believes of Jim's mother (b3) that she (b3) is clever, or

(S15b) Believes(Mike, Clever(b4)) & Mother-of(Jim, b3) & Equiv(b4, b3)

i.e., Mike believes of someone (b4) who is equivalent to Jim's mother (b3) that she (b4) is clever. Accordingly, CASSIE's representation of (*) might be as shown in Figure 8a or 8b. That is, roughly,

(*a) Believes(George, Believes(Mike, Clever(b3))) & Believes(George, Mother-of(Jim, b3))

or

(*b) Believes(George, Believes(Mike, Clever(b4))) & Believes(George, Mother-of(Jim, b3)) & Believes(George, Equiv(b4, b3))

Barnden says of CASSIE that she must represent (*) as in (*b),[10] which imputes a belief about Equiv to George, in violation of Strategy 1.

Our reply is quite simple: CASSIE would not represent (*) as in (*b) but as in (*a), which does not make the "harmful" imputation. But suppose that CASSIE's representations of (S15) and (*) were (S15b) and (*b): Is this "cognitively reasonable"? On the (b)-representations, CASSIE believes that George believes that there are two concepts: there is George's concept of Mike's concept of a clever person and there is George's concept of Jim's mother; and CASSIE believes that George believes that these two concepts are equivalent. But this equivalence is not an arcane imputation to George; it is (merely) *CASSIE's* belief.[11]

Barnden also raises two further objections (1986b: 346; 1989, Sect. 4): CASSIE has no way to quantify over intensions and no way to make statements about intensions themselves. For example, how would she represent these:

[10] Actually, he says, *de dicto*, that Maida and Shapiro (1982) must represent (*) thus; cf. Barnden 1986b: 345f.

[11] Arguably, (a) is less plausible than (b): On (a), CASSIE believes that George believes that there is a single concept: George's concept of Mike's concept of someone who *is*, on George's belief, Jim's mother. But surely it's plausible (and perhaps more so) for George to distinguish between his own concept of Jim's mother and Mike's concept of the clever person.

(3) Mike believes everything Bill believes.

(4a) Mike believes that some house he has a hazy idea of is red.

(4b) Mike believes that someone is a spy, but he has only a hazy idea of whom it might be.

The representation of (3) is simple; see Figure 9. Roughly, it is:

$$\forall v1[\text{Bill believes } v1 \rightarrow \text{Mike believes } v1]$$

The representations of (4a) and (4b) are a bit trickier. We must distinguish the concept of the house (or the spy) from the concept of the *idea of* the house (or the *idea of* the spy). (Cf. our earlier discussion of the difference between a concept and a concept of a concept.) We propose, tentatively, to represent this as in Figure 10. Then (4a) and (4b) would be represented as in Figures 11 and 12.[12]

7 CONCLUSIONS: BUILDING MINDS.

We have discussed knowledge representation—more accurately, we feel, the representation of beliefs—in the context of AI systems having what we call natural-language competence. Knowledge representation for natural-language competence is different from knowledge representation for robot or vision systems, because the goal of an NLC system is not to make its way in the real world, but to converse with humans about all the topics that humans discuss—real, imaginary, theoretical, and impossible.

We have also attempted to clarify, extend, and deepen our theory of intensional knowledge representation. The main points were that there is no need for multiple terms in the belief representation language of a cognitive agent for representing other cognitive agents in their roles as objects of belief and in their roles as agents of beliefs. Neither is there a need for multiple terms to represent propositions in their roles of propositions held by the cognitive agent, objects of normal beliefs, and objects of beliefs about believing them.

We are now comfortable in adopting what Searle 1980 refers to as the strong AI position: an Artificial Intelligence *is* a cognitive agent and *has* a mind. Perhaps a computer simulation of a hurricane is not a hurricane (cf. Rapaport 1988b), and certainly a model of a horse is not a horse. But a wind-tunnel simulation of air flow *is* air flow, a wind-tunnel model of an airfoil *is* an airfoil, a wave-tank model of a wave *is* a wave, a model of a statue of a horse *is* a statue of a horse, and a computer model of a mind (if successful) *is* a mind (cf. Rapaport 1986b). As John Haugeland says,

> Perhaps Artificial Intelligence should be called 'Synthetic Intelligence' to accord better with commercial parlance. Thus artificial diamonds are fake imitations, whereas synthetic diamonds are genuine diamonds, only manufactured instead of dug up .... Despite the name, AI clearly aims at genuine intelligence, not a fake imitation. (Haugeland 1985: 255.)

Thus, SNePS/CASSIE is both a model of a mind (some arbitrary, individual mind) and a mind itself (CASSIE's mind), and a node of SNePS/CASSIE is both a model of a concept as well as being (in Barnden's phrase) "something that is *used by* [CASSIE] as a basis for its normal cognitive processing" (Barnden 1986a: 411).

The task facing researchers in knowledge representation for NLC is to design a mind with the ability to handle the range and complexity of thoughts that humans have.

[12] The representation of (4b) is a bit more tentative than that of (4a), since it requires a theory of how to represent "knowing who": 'Mike has a (hazy) idea who the spy is' should probably be represented similarly to 'Mike knows who the spy is'.

8 ACKNOWLEDGMENTS.

This work was supported in part by the National Science Foundation under Grant IRI-8610517 (Rapaport and Shapiro), and in part by the Air Force Systems Command, Rome Air Development Center, Griffis Air Force Base, New York 13441-5700, and the Air Force Office of Scientific Research, Bolling AFB DC 20332 under Contract No. F30602-85-C-0008, which supports the Northeast Artificial Intelligence Consortium (NAIC) (Shapiro). The first author is grateful to Norman Sondheimer and the USC/Information Sciences Institute for their support and hospitality during his sabbatical year. Both authors are grateful to Soon Ae Chun for her work on CASSIE's grammar, and to the members of SNeRG (the SNePS Research Group), for their contributions and discussion. Earlier versions of portions of this essay appeared in Shapiro & Rapaport 1987b and Shapiro & Rapaport 1988.

9 REFERENCES.

1. Almeida, Michael J. (1987), "Reasoning about the Temporal Structure of Narratives," *Technical Report 87-10* (Buffalo: SUNY Buffalo Dept. of Computer Science).

2. Almeida, Michael J. & Shapiro, Stuart C. (1983), "Reasoning about the Temporal Structure of Narrative Texts," *Proceedings of the 5th Annual Conference of the Cognitive Science Society* (*University of Rochester*) (Hillsdale, NJ: Lawrence Erlbaum).

3. Anderson, J. R. (1977), "Memory for Information about Individuals," *Memory and Cognition* 5: 430-422.

4. Anderson, J. R. (1978), "The Processing of Referring Expressions within a Semantic Network," *Theoretical Issues in Natural Language Processing* (*TINLAP-2*) (New York: Association for Computing Machinery).

5. Asher, Nicholas (1986), "Belief in Discourse Representation Theory," *Journal of Philosophical Logic* 15: 127-89.

6. Asher, Nicholas (1987), "A Typology for Attitude Verbs and Their Anaphoric Properties," *Linguistics and Philosophy* 10: 125-197.

7. Barr, Avron, & Davidson, James (eds.) (1981), "Representation of Knowledge," in A. Barr & E. Feigenbaum (eds.), *The Handbook of Artificial Intelligence* (Los Altos, CA: Walter Kaufmann): 141-222.

8. Barnden, John A. (1986a), "A Viewpoint Distinction in the Representation of Propositional Attitudes," *Proceedings of the 5th National Conference on Artificial Intelligence* (*AAAI-86*; *Philadelphia)* (Los Altos, CA: Morgan Kaufmann), Vol. 1, pp. 411-415.

9. Barnden, John A. (1986b), "Imputations and Explications: Representational Problems in Treatments of Propositional Attitudes," *Cognitive Science* 10: 319-364.

10. Barnden, John A. (1989), "Towards a Paradigm Shift in Belief Representation Methodology," *Journal of Experimental and Theoretical Artificial Intelligence* 2: 133-161.

11. Brachman, Ronald J.; Fikes, Richard E.; & Levesque, Hector J. (1983), "Krypton: A Functional Approach to Knowledge Representation," *Computer* 16: 67-73; also in Brachman & Levesque 1985: 411-429.

12. Brachman, Ronald J., & Levesque, Hector J. (eds.) (1985), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kaufmann).

13. Brachman, Ronald J., & Schmolze, James G. (1985), "An Overview of the KL-ONE Knowledge Representation System," *Cognitive Science* 9: 171-216.

14. Brachman, Ronald J., & Smith, Brian C. (1980), "Special Issue on Knowledge Representation," *SIGART Newsletter*, No. 70.

15. Castañeda, Hector-Neri (1966), "'He': A Study in the Logic of Self-Consciousness," *Ratio* 8: 130-157.

16. Castañeda, Hector-Neri (1972), "Thinking and the Structure of the World," *Philosophia* 4 (1974) 3-40; reprinted in 1975 in *Critica* 6 (1972) 43-86.

17. Castañeda, Hector-Neri (1975a), "Identity and Sameness," *Philosophia* 5: 121-150.

18. Castañeda, Hector-Neri (1975b), *Thinking and Doing: The Philosophical Foundations of Institutions* (Dordrecht: D. Reidel).

19. Castañeda, Hector-Neri (1977), "Perception, Belief, and the Structure of Physical Objects and Consciousness," *Synthèse* 35: 285-351.

20. Castañeda, Hector-Neri (1984), "Philosophical Refutations," in J. H. Fetzer (ed.), *Principles of Philosophical Reasoning* (Totawa, NJ: Rowman & Allanheld): 227-258.

21. Castañeda, Hector-Neri (1989), "Philosophy as a Science and as a Worldview," in A. Cohen & M. Dascal (eds.), *The Institution of Philosophy* (Peru, IL: Open Court).

22. Cercone, Nick, & McCalla, Gordon (1987), (eds.), *The Knowledge Frontier*: *Essays in the Representation of Knowledge* (New York: Springer-Verlag).

23. Dennett, Daniel C. (1978), "Intentional Systems," in D. C. Dennett, *Brainstorms* (Montgomery, VT: Bradford Books): 3-22.

24. Dennett, Daniel C. (1982), "Beyond Belief," in A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press): xvi-95.

25. Fodor, Jerry A. (1975), *The Language of Thought* (New York: Thomas Y. Crowell Co.).

26. Frege, Gottlob (1892) "On Sense and Reference," trans. M. Black, in P. Geach & M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege* (Oxford: Basil Blackwell, 1970): 56-78.

27. Hardt, Shoshana L. (1987), "Conceptual Dependency," in Shapiro 1987: 194-199.

28. Haugeland, John (1985), *Artificial Intelligence*: *The Very Idea* (Cambridge, MA: MIT Press).

29. Hobbs, Jerry R. (1985), "Ontological Promiscuity," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics* (*University of Chicago*) (Morristown, NJ: Association for Computational Linguistics): 61-69.

30. Kamp, Hans (1984), "A Theory of Truth and Semantic Representation," in J. Groenendijk, T. M. V. Janssen, & M. Stokhof (eds.), *Truth*, *Interpretation and Information* (Dordrecht: Foris): 1-41.

31. Kamp, Hans (forthcoming), *Situations in Discourse*, manuscript (Austin, TX: University of Texas Center for Cognitive Science).

32. King, Margaret, & Rosner, Michael (eds.) (1986), "Special Issue on Knowledge Representation," *Proceedings of the IEEE* 74: 1299-1454.

33. Kramer, Bryan M., & Mylopoulos, John (1987), "Representation, Knowledge" in Shapiro 1987: 882-890.

34. Landman, Fred (1986), "Pegs and Alecs," in J. Y. Halpern (ed.), *Theoretical Aspects of Reasoning About Knowledge* (Los Altos, CA: Morgan Kaufmann): 45-61.

35. Levesque, Hector J. (1986), "Knowledge Representation and Reasoning," in J. F. Traub (ed.), *Annual Review of Computer Science*, Vol. 1 (Palo Alto: Annual Reviews): 255-287.

36. MacGregor, Robert, & Bates, Raymond (1987), "The Loom Knowledge Representation Language," *Proceedings of the Knowledge-Based Systems Workshop (St. Louis, MO)*; also, *Technical Report ISI/RS-87-188* (Marina del Rey, CA: USC/Information Sciences Institute).

37. Maida, Anthony S., & Shapiro, Stuart C. (1982), "Intensional Concepts in Propositional Semantic Networks," *Cognitive Science* 6: 291-330; reprinted in Brachman and Levesque 1985: 170-189.

38. Martins, João P., & Shapiro, Stuart C. (1983), "Reasoning in Multiple Belief Spaces," *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-83; Karlsruhe, W. Germany)* (Los Altos, CA: Morgan Kaufmann), Vol. 1, pp. 370-373.

39. Martins, João, & Shapiro, Stuart C. (1988), "A Model for Belief Revision," *Artificial Intelligence* 35: 25-79.

40. McKay, Donald P., & Shapiro, Stuart C. (1981), "Using Active Connection Graphs for Reasoning with Recursive Rules," *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI-81; Vancouver)*: 368-374.

41. Meinong, Alexius (1904), "Über Gegenstandstheorie," in R. Haller (ed.), *Alexius Meinong Gesamtausgabe*, Vol. II (Graz, Austria: Akademische Druck- u. Verlagsanstalt, 1971): 481-535; English translation ("The Theory of Objects") by I. Levi et al., in R. M. Chisholm (ed.), *Realism and the Background of Phenomenology* (New York: Free Press, 1960): 76-117.

42. Moser, M. G. (1983), "An Overview of NIKL, The New Implementation of KL-ONE," in C. Sidner, et al. (eds.), "Research in Knowledge Representation for Natural Language Understanding, Annual Report," *BBN Report No. 5421* (Cambridge, MA: Bolt Beranek and Newman).

43. Mylopoulos, John, & Levesque, Hector J. (1984), "An Overview of Knowledge Representation," in M. L. Brodie, J. Mylopoulos, & J. W. Schmidt (eds.), *On Conceptual Modeling* (New York: Springer-Verlag): 3-18.

44. Pollock, John (1988), "My Brother, The Machine," *Noûs* 22: 173-212.

45. Rapaport, William J. (1978), "Meinongian Theories and a Russellian Paradox," *Noûs* 12: 153-180; errata, *Noûs* 13 (1979) 125.

46. Rapaport, William J. (1981), "How to Make the World Fit Our Language: An Essay in Meinongian Semantics," *Grazer Philosophische Studien* 14: 1-21.

47. Rapaport, William J. (1985), "Meinongian Semantics for Propositional Semantic Networks," *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics (University of Chicago)* (Morristown, NJ: Association for Computational Linguistics): 43-48.

48. Rapaport, William J. (1985/1986), "Non-Existent Objects and Epistemological Ontology," *Grazer Philosophische Studien* 25/26: 61-95.

49. Rapaport, William J. (1986a), "Logical Foundations for Belief Representation," *Cognitive Science* 10: 371-422.

50. Rapaport, William J. (1986b), "Searle's Experiments with Thought," *Philosophy of Science* 53: 271-79.

51. Rapaport, William J. (1987), "Logic, Predicate," in Shapiro 1987: 538-544.

52. Rapaport, William J. (1988a), "Syntactic Semantics: Foundations of Computational Natural-Language Understanding," in J. Fetzer (ed.) *Aspects of Artificial Intelligence* (Dordrecht, Holland: Kluwer Academic Publishers): 81-131.

53. Rapaport, William J. (1988b), "To Think or Not to Think," *Noûs* 22: 585-609.

54. Rapaport, William J.; Segal, Erwin M.; Shapiro, Stuart C.; Zubin, David A.; Bruder, Gail A.; Duchan, Judith F.; & Mark, David M. (1989), "Cognitive and Computer Systems for Understanding Narrative Text," *Technical Report 89-07* (Buffalo: SUNY Buffalo Dept. of Computer Science, August 1989).

55. Rapaport, William J., & Shapiro, Stuart C. (1984), "Quasi-Indexical Reference in Propositional Semantic Networks," *Proceedings of the 10th International Conference on Computational Linguistics* (*COLING-84*; *Stanford University*) (Morristown, NJ: Association for Computational Linguistics): 65-70.

56. Rapaport, William J.; Shapiro, Stuart C.; & Wiebe, Janyce M. (1986), "Quasi-Indicators, Knowledge Reports, and Discourse," *Technical Report 86-15* (Buffalo: SUNY Buffalo Department of Computer Science).

57. Russell, Bertrand (1905), "On Denoting," reprinted in *Logic and Knowledge* (ed. R. C. Marsh) (New York: Capricorn, 1956): 39-56.

58. Schank, Roger C. (1975), *Conceptual Information Processing* (New York: Elsevier).

59. Schank, Roger C., & Rieger, Charles J. (1974), "Inference and the Computer Understanding of Natural Language," *Artificial Intelligence* 5: 373-412; reprinted in Brachman and Levesque 1985: 120-139.

60. Schank, Roger C., & Riesbeck, Christopher K. (eds.) (1981), *Inside Computer Understanding*: *Five Programs Plus Miniatures* (Hillsdale, NJ: Lawrence Erlbaum).

61. Searle, John R. (1980), "Minds, Brains, and Programs," *Behavioral and Brain Sciences* 3: 417-457.

62. Sells, Peter (1987), "Aspects of Logophoricity," *Linguistic Inquiry* 18: 445-479.

63. Shapiro, Stuart C. (1978), "Path-Based and Node-Based Inference in Semantic Networks," in D. Waltz (ed.), *Theoretical Issues in Natural Language Processing* (*TINLAP-2*) (New York: Association for Computing Machinery): 219-225.

64. Shapiro, Stuart C. (1979), "The SNePS Semantic Network Processing System," in N. V. Findler (ed.), *Associative Networks*: *The Representation and Use of Knowledge by Computers* (New York: Academic Press): 179-203.

65. Shapiro, Stuart C. (1982), "Generalized Augmented Transition Network Grammars for Generation from Semantic Networks," *American Journal of Computational Linguistics* 8: 12-25.

66. Shapiro, Stuart C. (ed.) (1987), *Encyclopedia of Artificial Intelligence* (New York: John Wiley & Sons).

67. Shapiro, Stuart C.; Martins, João P.; & McKay, Donald P. (1982), "Bi-Directional Inference," *Proceedings of the 4th Annual Conference of the Cognitive Science Society* (Hillsdale, NJ: Lawrence Erlbaum): 90-93.

68. Shapiro, Stuart C., & McKay, Donald P. (1980), "Inference with Recursive Rules," *Proceedings of the 1st National Conference on Artificial Intelligence* (*AAAI-80*; *Stanford University*) (Los Altos, CA: Morgan Kaufmann), Vol. 1: 151-153.

69. Shapiro, Stuart C., & Rapaport, William J. (1987a), "SNePS Considered as a Fully Intensional Propositional Semantic Network," in Cercone & McCalla 1987: 262-315; earlier version preprinted as *Technical Report No. 85-15* (Buffalo: SUNY Buffalo Department of Computer Science, 1985); shorter version appeared in *Proceedings of the Fifth National Conference on Artificial Intelligence* (*AAAI-86*, *Philadelphia*) (Los Altos, CA: Morgan Kaufmann): 278-283.

70. Shapiro, Stuart C., & Rapaport, William J. (1987b), "Knowledge Representation for Natural Language Processing," *Presentations from the 1987 Natural Language Planning Workshop*, *Northeast Artificial Intelligence Consortium Technical Report WR-8703* (Griffis Air Force Base, NY: Rome Air Development Center): 56-77.

71. Shapiro, Stuart C., & Rapaport, William J. (1988), "Models and Minds: A Reply to Barnden," *Northeast Artificial Intelligence Consortium Technical Report TR-8737* (Syracuse: Syracuse University, March 1988); reprinted as "Models and Minds", *Computer Science Research Review* (Buffalo: SUNY Buffalo Dept. of Computer Science, 1988-1989): 24-30.

72. Smith, Brian C. (1982), "Reflection and Semantics in a Procedural Language," *Technical Report MIT/LCS/TR-272* (Cambridge, MA: MIT Laboratory for Computer Science); page reference is to the prologue, reprinted in Brachman & Levesque 1985: 32-39.

73. Stich, Stephen (1983), *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge, MA: MIT Press).

74. Vere, Steven A. (1987), "Planning," in Shapiro 1987: 748-758.

75. Vilain, Marc (1985), "The Restricted Language Architecture of a Hybrid Representation System," *Proceedings of the 9th International Joint Conference on Artificial Intelligence* (*IJCAI-85*; *Los Angeles*) (Los Altos, CA: Morgan Kaufmann): 547-551.

76. Weischedel, Ralph M. (1986), "Knowledge Representation and Natural Language Processing," *Proceedings of the IEEE* 74: 905-920.

77. Wiebe, Janyce M., & Rapaport, William J. (1986), "Representing *De Re* and *De Dicto* Belief Reports in Discourse and Narrative," *Proceedings of the IEEE* 74: 1405-1413.

78. Wilks, Yorick (1986), "Relevance and Beliefs," in T. Myers, K. Brown, & B. McGonigle (eds.), *Reasoning and Discourse Processes* (London: Academic Press): 266-289; earlier version preprinted as "Relevance, Points of View and Speech Acts: An Artificial Intelligence View," *Memoranda in Computer and Cognitive Science No. MCCS-85-25* (Las Cruces: New Mexico State University Computing Research Laboratory).

79. Winograd, Terry (1975), "Frame Representations and the Declarative/Procedural Controversy," in D. G. Bobrow & A. M. Collins (eds.), *Representation and Understanding*: *Studies in Cognitive Science* (New York: Academic Press): 185-210; reprinted in Brachman & Levesque 1985: 358-370.

80. Wyatt, Richard W. (1989), "The Representation of Opaque Contexts," *Technical Report* (Buffalo: SUNY Buffalo Dept. of Computer Science, forthcoming).