What Did You Mean by That? Misunderstanding, Negotiation and Syntactic Semantics

William J. Rapaport

Department of Computer Science and Engineering,
Department of Philosophy, and Center for Cognitive Science
State University of New York at Buffalo, Buffalo, NY 14260-2000

rapaport@cse.buffalo.edu
http://www.cse.buffalo.edu/∼rapaport/

December 17, 2002

Abstract

Syntactic semantics is a holistic, conceptual-role-semantic theory of how computers can think. But Fodor & Lepore have mounted a sustained attack on holistic semantic theories. However, their major problem with holism (that, if holism is true, then no two people can understand each other) can be fixed by means of *negotiating meanings*. Syntactic semantics and Fodor & Lepore's objections to holism are outlined; the nature of communication, miscommunication, and negotiation is discussed; Bruner's ideas about the negotiation of meaning are explored; and some observations on a problem for knowledge representation in AI raised by Winston are presented.

1 Introduction

When you and I speak or write to each other, the most we can hope for is a sort of incremental approach toward agreement, toward communication, toward common usage of terms. (Lenat 1995: 45.)

'Syntactic semantics' is the name I have given to a theory of how computers can think (Rapaport 1986, 1988, 1995, 1996, 1998, 1999, 2000b, 2002). Syntactic semantics is a kind of conceptual-role semantics and (therefore) holistic in nature (Rapaport 2002). But Jerry Fodor and Ernest Lepore (1991, 1992) have mounted a sustained attack on holistic semantic theories. The present essay is a sustained reply to their main complaint. (I gave a shorter reply in Rapaport 2002.) Briefly, I believe that Fodor & Lepore's major problem with holism—that, if holism is true, then no two people can understand each other—can be fixed by means of *negotiating meanings*. In what follows, I outline syntactic semantics and Fodor & Lepore's objections to holism; discuss the nature of communication, miscommunication, and negotiation; explore some of Jerome Bruner's ideas about the negotiation of meaning; and conclude with some observations on a problem for knowledge representation raised by Patrick Henry Winston (and possibly by Wittgenstein).

2 Background: AI as Computational Cognition

'Syntactic semantics' is my name for the theory that underlies virtually all AI research, or at least that portion of it that falls under the rubrics of "computational psychology" and "computational philosophy" (cf. Shapiro 1992).

The goal of *computational psychology* is to study human cognition using computational techniques. Computational-psychology theories are expressed (i.e., implemented)[1] in computer programs. Good computational-psychology programs will (when executed)[2] replicate human cognitive tasks in ways that are faithful to human performance, with the same failures as well as successes—AI as computational psychology can tell us something about the human mind.

The goal of *computational philosophy* is to learn which aspects of cognition in general are computable. Good computational-philosophy computer programs

[1]For relevant discussion of the nature of implementation, see Rapaport 1996, Ch. 7, and Rapaport 1999.

[2]On the important distinction between a (textual) *program* and the *process* created when the program is executed, see Rapaport 1988, §2.

will (when executed) replicate cognitive tasks but not necessarily in the way that humans would do them. AI as computational philosophy can tell us something about the scope and limits of cognition in general (e.g., which aspects of it are computable), but not necessarily about human cognition in particular (Rapaport 1998, cf. Rapaport 2000a).

Together, we can call these "computational cognition", a term I prefer over 'artificial intelligence', since AI is really less concerned with "intelligence" in particular (whatever that might be), and more with cognition in general (human or otherwise). I think it fair to say that many AI researchers, when trying to explain what AI is, have to say, somewhat defensively, that by 'intelligence' they *don't* (normally) mean intelligence in the sense of IQ. We are not (normally) trying to build high-IQ computers (although that might be a considerably easier task!). Rather, we really mean by the term 'intelligence' something more like cognition in general: the gathering, storing, processing, and communicating of information (cf. Győri 2002: 133-134, citing Neisser 1976 and Geeraerts 1997). And there is nothing "artificial" about it: Cognition modeled computationally is real cognition, not just a simulation of it (Rapaport 1988, 1998).[3]

3 Syntactic Semantics: An Outline

Syntactic semantics has three basic theses (which I have argued for in detail in the essays cited earlier):

3.1 Semantics is "internal" (or "psychological"), and therefore syntactic

Consider a single domain of uninterpreted "markers" (or "symbols"). *Syntax* is classically considered to be the study of the relations *among* these markers, i.e., the study of the rules of "symbol" formation and manipulation (Morris 1938: 6–7; cf. Posner 1992). Call this domain the 'syntactic domain'. Often, it is called a 'formal system', but I wish to consider other "systems", including physical ones, such

[3]There is a third kind of AI research, which is not relevant to the present concerns: AI as "advanced computer science". Its goal is to solve general problems in any area of computer science by applying the methods of AI, but it is not directly concerned with cognitive issues (cf. Shapiro 1992). It is at the "cutting edge" of computer science, as my former colleague John Case once put it. (Another former colleague, Anthony Ralston, used a topologically equivalent, but opposite, metaphor: He told me that AI is at the "periphery" of computer science!)

as the brain considered as a neural network. Also, I prefer to talk about "markers" rather than the more usual "symbols", since I want to think of the markers independently of their representing (or not representing) things external to the syntactic domain, and I don't want to rule out connectionist theories even though I favor "good old-fashioned, classical, symbolic AI" theories (see Rapaport 2000b for further discussion).

By contrast, *semantics* is classically considered to be the study of relations *between two* domains: the uninterpreted markers of the syntactic domain and interpretations of them in a "semantic domain" (again, cf. Morris 1938: 6–7, Posner 1992).

But, by considering the *union* of the syntactic and semantic domains, semantics (classically understood) can be turned into syntax (classically understood): i.e., semantics can be turned into a study of relations within a single domain among the markers and their interpretations. This is done by incorporating (or "internalizing") the semantic interpretations along with the markers to form a unified system of *new* markers, some of which are the old ones and the others of which are their interpretations.

Hence, syntax (i.e., "symbol" manipulation of the new markers) *can* suffice for the semantical enterprise, *pace* one of the explicit premises of John Searle's Chinese-Room Argument (1980). In particular, syntax suffices for the semantics needed for a computational cognitive theory of natural-language understanding and generation (Rapaport 2000b).

This is precisely the situation with respect to "conceptual" and "cognitive" semantics, in which both the linguistic expressions we use and their meanings-as-mental-concepts are located "in the head". As I once said a long time ago (Rapaport 1981), we must make the "world" fit our language, or, as Ray Jackendoff has more recently put it, we must "push 'the world' into the mind" (Jackendoff 2002, §10.4). It also appears to be the situation with contemporary Chomskian semantics.[4]

This is also the situation with respect to our brains: Both our "mental" concepts as well as the output of our perceptual processes (i.e., our internal representations of external stimuli) are "implemented" in a single domain of neuron firings, some of which correspond to (e.g., are caused by) things in the external

[4]On cognitive semantics, cf. Lakoff 1987, Talmy 2000. On conceptual semantics, cf. Gärdenfors 1997, 1999ab; and Jackendoff 2002, esp. Ch. 9 ("Semantics as a Mentalistic Enterprise"), Ch. 10 ("Reference and Truth"), and, most especially, Jackendoff 2002, Ch. 10, §10.4 ("Pushing 'the World' into the Mind"). On Chomskian semantics, cf. McGilvray 1998 and discussion in Rapaport 2000b.

world and some of which are our "mental" concepts of those things.

3.2 Semantics is recursive, and therefore *at bottom* syntactic

Semantics can also be considered as the process of understanding the syntactic domain (by modeling it) in terms of the semantic domain. But if we are to understand one thing in terms of another, that other thing must be antecedently understood. Hence, semantics on this view is recursive: The semantic domain can be treated as a (new) syntactic domain requiring a further semantic domain to understand *it*, in what Brian Cantwell Smith (1987) has called a "correspondence continuum". To prevent an infinite regress, some domain must be understood in terms of itself. (Or some cycle of domains must be understood in terms of each other; see Rapaport 1995.) This base case of semantic understanding can be called "syntactic understanding": understanding a (syntactic) domain by being conversant with manipulating its markers, as when we understand a deductive system proof-theoretically rather than model-theoretically, or as when we understand how to solve high-school algebra problems by means of a set of syntactic rules for manipulating the variables and constants of an equation rather than by means of a set of semantic rules for "balancing" the equation (see Rapaport 1986 for details).

3.3 Syntactic semantics is methodologically solipsistic

The internalization of meanings ("pushing the world into the mind", to use Jackendoff's phrase) leads to a "narrow" or first-person perspective on cognition. Moreover, this point of view is all that is needed for understanding or modeling cognition: An "external", or "wide", or third-person point of view may shed light on the nature of correspondences between cognition and the external world, but it is otiose[5] for the task of understanding or modeling cognition. (For discussion, see Maida & Shapiro 1982; Rapaport 1985/1986, 2000b; Shapiro & Rapaport 1991; Rapaport, Shapiro, & Wiebe 1997.)

[5] A wonderfully non-otiose (i.e., useful) word meaning "of no use; ineffective; futile; serving no useful purpose; having no excuse for being"; cf. [http://www.dictionary.com/search?q=otiose].

4 Syntactic Semantics and Holism

> *Researchers concerned with modeling people recognize that people cannot be assumed to ever attribute precisely identical semantics to a language. However, the counterargument is that computers can be programmed to have precisely identical semantics (so long as they cannot modify themselves). Moreover, as evidenced in human coordination, identical semantics is not critical, so long as satisfactory coordination can arise.* (Durfee 1992: 859.)

Syntactic semantics as I just sketched it is a holistic conceptual-role semantics that takes the meaning of an expression for a cognitive agent to be that expression's "location" in the cognitive agent's semantic network of (all of) the cognitive agent's other expressions (cf. Quillian 1967, 1968, 1969; Rapaport 2002). But, according to Fodor & Lepore, holistic semantic theories are committed to the following "*prima facie* outlandish claims" (1991: 331, emphasis theirs):

> that no two people ever share a belief; that there is no such relation as translation; that no two people ever mean the same thing by what they say; that no two time slices of *the same* person ever mean the same thing by what they say; that no one can ever change his [sic] mind; that no statements, or beliefs, can ever be contradicted . . . ; and so forth.

The third claim is central: *No two people ever mean the same thing by what they say.* It is central because, along with reasonable cognitivist assumptions about the relations of meaning to beliefs and to language, it can be used to imply all the others. I think the third claim is true, but only *prima facie* outlandish. So did Bertrand Russell (1918: 195–196), who also thought it necessary for communication:

> When one person uses a word, he [sic] does not mean by it the same thing as another person means by it. I have often heard it said that that is a misfortune. That is a mistake. It would be absolutely fatal if people meant the same things by their words. It would make all intercourse impossible, and language the most hopeless and useless thing imaginable, because the meaning you attach to your words must depend on the nature of the objects you are acquainted with, and since different people are acquainted with different objects, they would not be able to talk to each other unless they attached quite different meanings to their words. . . . Take, for example, the word 'Piccadilly'. We, who are acquainted with Piccadilly, attach quite a different meaning to that word from any which could be attached to it by a person who had never been in London: and, supposing that you travel in foreign parts and expatiate on Piccadilly, you will convey to your hearers entirely

> different propositions from those in your mind. They will know Piccadilly as an important street in London; they may know a lot about it, but they will not know just the things one knows when one is walking along it. If you were to insist on language which was unambiguous, you would be unable to tell people at home what you had seen in foreign parts. It would be altogether incredibly inconvenient to have an unambiguous language, and therefore mercifully we have not got one.

So, if Fodor & Lepore's third claim is true, then how is it that we can—apparently successfully—communicate? The answer, I believe, can be found in negotiation. Let me briefly sketch how.

5 The Paradox of Communication

Since syntactic semantics is methodologically solipsistic, it would appear to be isolated—insulated, if you will—from the rest of the world of language and communication. But surely that cannot be the case. The external world *does*, of course, impinge on the internal syntactic system.

First, we have internal representations of external stimuli that cause their internal "representatives". Not all (internal) thoughts are thoughts of external things, of course; we can, famously, have hallucinations, phantom-limb pain (cf. Melzack 1992), and thoughts of unicorns and Santa Claus. And some of our internal thoughts of external things might not be *caused* by their external counterparts; we can, for example, have thoughts—caused (or derived in some way) by *other* thoughts—that only later are observed to have external counterparts (e.g., black holes "existed" as theoretical entities before they were observed).

Second, we *do* communicate with others. This, after all, is how we avoid *real* first-person solipsism, as opposed to the merely methodological kind. And when we communicate, it can be about an internal thing (as when I tell you a story or describe my hallucination) or about an external thing (as when I warn you of a large bee hovering around you). Of course, to be more precise, in the latter case, I am really describing my internal representation of the external bee, since that is all that I am directly aware of (cf. Rapaport 2000b and §§6–8, below). However, if there is an external bee that is a counterpart of my internal one that I am describing, you had better beware.

Now, in order for two cognitive agents to communicate *successfully*, whether about an internal thing or an external thing, they must be able to detect misunderstandings and correct them by negotiating: When we communicate, we attempt to convey our internal meanings to an audience (a hearer or reader)[6] by means of a "public communication language" (I owe this term to Shapiro 1993). You do not have direct access to my thoughts, but only to my speech acts (more generally, to my "language acts", including both speech and writing, as well as gestures).[7] You don't interpret what I am privately thinking; you can only interpret my public language and gestures. Your interpretation is, in fact, a *conclusion*; understanding involves inference, albeit defeasible inference. And it is not only occasionally defeasible: We almost *always* fail (this is Fodor & Lepore's third claim). Yet we almost always nearly succeed: This is the paradox of communication. Its resolution is simple: Misunderstandings, *if small enough*, can be ignored. And those that cannot be ignored can be minimized through negotiation. In this way, we learn what our audience meant or thought that we meant. Children do that when learning their first language (Bruner 1983; see §8, below). Adults continue to do it, which suggests that one of the processes involved in first-language acquisition continues to be involved in adult language use. (Indeed, some linguists claim that misinterpretation can lead to semantic change; cf. Győri 2002, §§1,5, citing, *inter alia*, Traugott 1999.) Thus, the same processes seem to be involved in all stages of the development of (a) language.

6 Communication.

> *They'd entered the common life of words. … After all, hadn't the author of this book turned his thoughts into words, in the act of writing it, knowing his readers would decode them as they read, making thoughts of them again?* (Barker 1987: 367.)

Consider two cognitive agents, Cassie and Oscar (either computer or human), who are interacting using a public communication language, e.g., English. Clearly, they need to be able to understand each other. If Cassie uses an expression in a way not immediately understood by Oscar, then Oscar must be able to clarify the situation (perhaps to correct the misunderstanding) *using the public communication language,* and not, say, by reprogramming Cassie (cf. Rapaport 1995, §2.5.1; we'll return to this idea in the next section). To do this, Oscar must find out, via language, what Cassie intended ("meant") to say. So, they must communicate.

[6]Wherever I write about "speaker" and "hearer", I normally also mean to include "writer" and "reader", and vice versa.

[7]The term 'language act' is due to Kearns, forthcoming 2003, and includes thinking with words as well as speech and writing. On gestures, cf. Gracia 1990: 495; see §9, below.

The theory of syntactic semantics places a fairly heavy burden on the role of communication. This may, at first glance, seem odd for a theory that claims to take a first-person, methodologically solipsistic point of view. But, as we will see (§8), all that matters is each interlocutor's perspective on the conversation, not any "global" perspective.

As expressed in the quotation that opens this section, a standard way of looking at communication is that the only way for me to know what's going on in your mind is for you to express your ideas in language—to "implement" them in words—and for me to translate from that public communication language into my own ideas. The public communication language thus plays much the same role as an interlingua does in machine translation (although "inside out": the public communication language is intermediate between two mental languages of thought, whereas a machine-translation interlingua is very much like a language of thought that is intermediate between two public communication languages; for more information on interlinguas, see, e.g., Slocum 1985: 4).

When we read, we seemingly just stare at a bunch of arcane marks on paper, yet we thereby magically come to know of events elsewhere in (or out!) of space and time.[8] How? By having an algorithm that maps the marks (which have a syntax) into our concepts, i.e., by interpreting the marks.[9] Conversely, in speaking, my ideas travel from my mind, to my mouth, to your ears, to your mind. This image is commonplace in both popular culture and academic discourse: Consider a "Garfield" Post-It Note that my colleague Stuart C. Shapiro once sent me, with the caption "From my mind to yours", showing a dashed line emanating from Garfield-the-cat's head to Odie-the-dog's head, which lights up in response; here, information gets sent from the sender's (i.e., Garfield's) mind to the recipient's (i.e., Odie's) by being written on paper (i.e., by being implemented in language). Or consider an illustration from Saussure 1959 showing two heads, labeled "A" and "B", with a line emanating from A's head through his mouth to B's ear and then to B's head, thence from B's head through B's mouth to A's ear and then

[8] I owe this ancient observation to someone else, but I can no longer recall whom! Possibly Clifton Fadiman?

[9] Here, I am viewing reading in *accord* with the view of at least one standard textbook on reading education (Harris & Sipay 1990: 1)—"Reading is a complex process. In some manner yet to be fully understood, the reader combines information provided by an author via printed or handwritten text with previously possessed knowledge to construct an interpretation of that text" (cf. Harris & Sipay 1990: 10)—and *differently* from Wittgenstein (1958, §§156–171), who does "not count . . . the understanding of what is read as part of 'reading' " (with the caveat, "for purposes of this investigation").

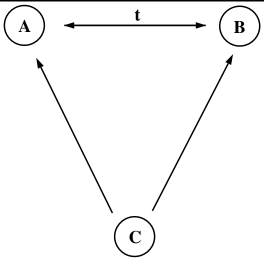


Figure 1: *A* and *B* are cognitive agents communicating about *C*, a real object in the external world. The arrow between *A* and *B* represents the communication between them of some term *t* expressing *C*. The arrows from *C* to *A* and to *B* represent *A*'s and *B*'s (joint) sensory access to *C*.

returning to A's head.

When we communicate in this way, what are we communicating about? It would seem, from the simplified Garfield and Saussure images, that we are only talking about our own ideas. What about the real world? Surely, we often talk about some external object that the two of us have joint access to (suppose I do want to warn you about the hovering bee). Isn't that, after all, how we know that we're talking about the same thing? Isn't the picture really as in Figure 1? In Figure 1, the idea is that two cognitive agents *A* and *B* use some term *t* of a public communication language to refer to some external object *C* in the real world. Both *A* and *B* have independent, direct access to *C*, and so can adjust their understanding of *t* by comparing what the other says about *C* with *C* itself. Is that not how things work?

As is often the case, the answer is: Yes and No. The picture is too simple. One missing item is that *A*'s access to *C* results in a private, internal idea (or set of ideas) about *C*, and similarly for *B* (see Rapaport 2000b for discussion). On the first-person point of view, *it is these private, internal ideas* that *A* and *B* are talking about (i.e., trying to communicate something about), *not C*. If we merge Saussure's image with ours, we get Figure 2 (in the spirit of Rube Goldberg).[10] Here, cognitive agent *A* perceives external object *C* and constructs (or retrieves) her own mental representation of *C*; call it C_A. Cognitive agent *A* then wishes to inform cognitive agent *B* of what she (*A*) is thinking, and so utters *t*, some expres-

[10]For readers unfamiliar with Rube Goldberg's cartoons, see Wolfe 2000 or go to [http://www.rube-goldberg.com/].

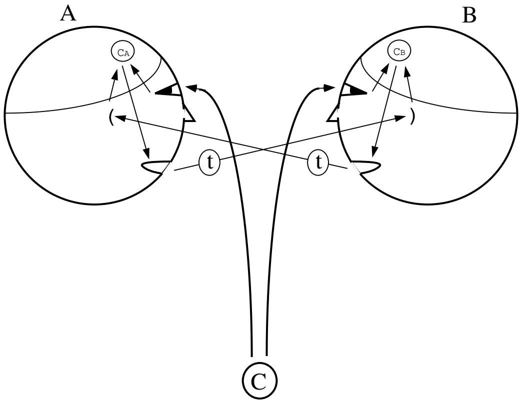


Figure 2: A perceives C, constructs or retrieves mental representation C_A of C, and utters t; B hears t and constructs or retrieves mental representation C_B. Similarly, B perceives C, constructs or retrieves C_B, and utters t; A hears t and constructs or retrieves C_A. (See text.)

sion of a public communication language that Fregeanly denotes C and internally means C_A. Cognitive agent B hears t and constructs (or retrieves) his own mental representation, C_B.

The arrows in Figure 2 from C to A's eyes and to B's represent the causal connections between C and A and between C and B. The arrows from A's eyes and ears to C_A and from B's to C_B represent the retrieval or production of their individualized, private, "perspectival" objects C_A and C_B. The arrows labeled 't' from A's mouth to B's ear (and from B's mouth to A's ear) represent the attempt at communication using some term t.

The public linguistic expression t is "symbolic currency" used "to exchange meaning" (Sacks 1989, quoted in Sacks 1990: 3). Expressions like t constitute the text of the "books" that enable us "to be with another's mind". Actually, speech-recognition problems can produce even more distance between what A and B are trying to communicate, for B's perception of t might differ from what A intended to produce.[11] For example, if A speaks English with a certain foreign accent, then the word that A intended to utter as 'seller' might be heard by B as 'sailor', or 'b's and 'p's (or 'l's and 'r's) might be confused, or, closer to home, it might not be immediately clear if the speaker said 'cereal' or 'serial'.[12] (See §9 for yet another

[11] As my colleague J.P. Koenig reminded me.

[12] When I was a graduate student at Indiana University in the early 1970s, my fellow students

caveat.)

But what does t mean? What gets communicated? Note that C_A and C_B are the psychological meanings of t that Frege repudiated but that conceptual and cognitive linguists embrace. They can be thought of as (representing) Meinongian objects or Castañedian guises or propositions (see, e.g., Meinong 1904, Castañeda 1972, Rapaport 1978). Note further that Frege's "sense" does not seem to play a role in communication understood in this perspectival way, despite (or perhaps due to!) its intersubjective or "social" nature. (On social language, cf. Gärdenfors 1993.) However, if we focus on two of Frege's principles about senses (viz., that every expression has one, and that an expression's sense is the way that expression refers to its referent (if it has one)) and if we parameterize these principles to an *individual's* use of a word, then C_A is indeed A's sense of t (and C_B is B's sense of t).

What role, then, does poor C—the actual object "out there" in the world—play? Very little, other than being there and causally producing C_A. In fact, on the first-person point of view, A and B will typically merely *hypothesize* the existence of C to the extent that their communicative negotiations are constrained by the roles that C_A and C_B play in their respective knowledge bases. A can only access C indirectly via C_A (and B can only access it indirectly via C_B).

Suppose A talks about what A thinks of as C_A, namely, A's own mental representation of C. There are then two possibilities:

1. B takes A to be talking about what B thinks of as C_B, namely, B's own mental representation of C; i.e., B understands A.

2. B takes A to be talking about something distinct from C_B; i.e., B misunderstands A.

Case 1 is close to the ideal situation, the case of "perfect" mutual understanding. In this case, B comes to believe that A is thinking of the "same" thing that he (B) is thinking of. B could continue the conversation by saying something else about C_B, using t. A hears t and constructs (or retrieves) her mental representation in one of two ways, just as B did. Again, then, we have either a case-1 of perfect understanding or a case-2 of misunderstanding.

---

and I heard Richard Routley give a lecture about endangered species, using what we took to be a thought-experiment about a creature called the "blue wile". It wasn't till his talk was over that we realized he wasn't making it up, but had all along been discussing—in his Australian accent—the blue *whale* (pronounced /wale/ by us Yanks).

Case 2 is the case of miscommunication, of misunderstanding. In this case, where *B* has misunderstood *A*, *B* might (eventually) say something that will alert *A* to the misunderstanding. *B* might, for instance, say something that makes no sense to *A*, and so *A* will realize that *B* misunderstood. Or *B* might hear *A* say things that make no sense to *B*, and so *B* will realize that he (*B*) misunderstood and alert *A* to this fact. By continued communication, *A* and *B* will "negotiate" about what it is they are talking about, hopefully eventually coming to some agreement.

7 Negotiation.

Miscommunication (case 2, above) is in fact the norm. Suppose that you perceive *C*, which causes you to think of C_{you}, which, in turn, leads you to utter some term t. And suppose that, when I hear your utterance of t, I think of C_I. Even if this C_I is the C_I that I think of when I *perceive* *C*, still, C_I will play a different role in my network of concepts and beliefs than C_{you} does in yours, simply because my conceptual network will be different from yours. As Fodor & Lepore said, semantic holism implies that no two people ever mean the same thing by what they say. So how *do* we successfully communicate and understand each other, as—apparently, or for all practical purposes—we do?

> Interpretations are negotiated in interaction. Every time we talk, we negotiate interpretations about referential and social meanings. The more intense and frequent the interaction between speakers with diverging interpretations of the meanings of a word, the more likely a 'negotiated settlement' will obtain, more or less spontaneously, through linguistic usage. When interpretations become conventionalized, we call that 'meaning'. But even so, that new meaning is subject to revision and negotiation. (Alvarez 1990.)

Candace Sidner (1994) points out that discourses among collaborators function as negotiations and that discourses containing negotiations serve to establish mutual beliefs. Negotiation is the key to understanding.

Negotiation can take many forms. Communicative negotiation plays a role when a cognitive agent understands by *translating* (Rapaport 2002, §6.6). If a translation seems not to preserve truth (i.e., seems to be inconsistent with the agent's understanding of the world), then negotiation with the interlocutor can bring understanding by restoring consistency. (And in this way Fodor & Lepore's *second* "outlandish claim" can be avoided.)

A form of negotiation also plays a role when a cognitive agent is *reading* and comes across an unknown word that is not in any available dictionary or there is no one the reader can ask who knows its meaning. In such a case, the reader can hypothesize a meaning from the context (including the reader's background knowledge), and revise that meaning as needed when the word is seen again. This is a sort of negotiation, not with a live interlocutor, but directly with the text (see Ehrlich 1995, Rapaport & Ehrlich 2000, Rapaport & Kibby 2002).

It is through the process of interactively (i.e., reciprocally) communicating with others that cognitive agents come to learn language. Here, negotiation can take the form of self-organization: "Given a system in which there is natural variation through local fluctuations [read: individual differences in meanings], global coherence ... may emerge provided certain kinds of positive feedback loops [read: negotiation] are in place" (Steels 1998: 138).

Negotiation can also correct misunderstandings, and facilitate us in changing our minds (*pace* Fodor & Lepore's fifth "outlandish claim"). Such communication allows one to "align" one's own knowledge base, expressed in one's own language of thought, with another's. It can also enable one computational agent to align its ontology with another's (cf. Campbell & Shapiro 1998, Campbell 1999). In short, communicative negotiation can resolve conflicts, enabling us to understand one another.

Perception can play a role in negotiation.[13] It, too, is a kind of "communication" with something external to the understander. Crucially, both perception and communication work in the same way: The understander compares two *internal* representations: one is causally produced from the speaker or the act of perception; the other is part of the antecedently-existing internal mental network. When there is a mismatch, the understander must change his or her (or its) mind. "As long as the conversation proceeds without our getting into ... [a] situation" in which "we *didn't* know what was meant", the cognitive agent "has all the connections with reality it needs" (Shapiro & Rapaport 1987: 271).

But why does the problem arise at all? Why is there the potential for (and usually the actuality of) miscommunication resulting in misunderstanding? The answer is simple:

> ... transmission of representations themselves is impossible. I cannot be sure that the meaning of a word I say is the same for the person to whom I direct it. Consequently, language works as a system of values, of reciprocal expectations. To say it differently, the processes of verbal communication always constitute a try, a hypothesis, and an intention from the sender to the

[13] Cf. Maida & Shapiro 1982: 300–301 on the usefulness of sensors and effectors for this purpose. And see Shapiro & Ismail 2001 for a discussion of ways of implementing this.

receiver. (Vauclair 1990: 321–322.)

On this view, if you could literally read my mind (as, indeed, I as programmer *can* literally read the "mind" of my computational cognitive agent Cassie), there would be no misunderstanding, hence no miscommunication. But, since you can't, there is.

This inability to read minds is likewise a commonplace of popular culture: It is illustrated nicely by a *Cathy* cartoon in which the following dialogue ensues while Cathy and her boyfriend Irving are sitting on a couch, with Cathy's dog between them.

> Irving (looking at the newspaper): We could still go out.
> Cathy (looking grim): It's too late.
> Irving: Too late for what?
> Cathy: I wanted to go out before. I don't want to go out now.
> Irving: If you wanted to go out, why didn't you say so, Cathy?
> Cathy: I wanted you to say you wanted to, Irving.
> Irving: I said we could go if you wanted to.
> Cathy: Why would I want to if you didn't act as if you wanted to?
> Irving: Why didn't you just *say* you wanted to?
> Cathy: I said I wanted to by getting mad when you didn't say you wanted to.
> Irving: I said I wanted to by saying I would.
> Cathy: Why didn't you just say what you meant?
> Irving: Why didn't *you* say what you meant?
> Cathy: Hah!!
> Irving and Cathy (together): *My words came out fine! They were processed incorrectly by your brain!!!*
> Cathy's dog (thinking): Few things are as scary as being between two humans who are agreeing with each other.

Carnap has argued that the "mind-reading" method, which he calls "structure analysis", is superior to the "behavioristic" method that we are in fact restricted to (Carnap 1956: 244–247, on "The Concept of Intension for a Robot"; cf. Simon 1992, esp. pp. 6–7, for discussion and elaboration of Carnap's "structural" method and its application to the Chinese-Room Argument(!)). Arguably, though, even such literal mind-reading wouldn't suffice. For you would still have to understand my language of thought, just as a reader of a text written in one's native language must interpret that text even though the language is common. So it's highly unlikely, except possibly in the most artificial of situations (as with a computational

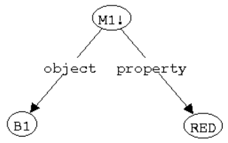


Figure 3: A SNePS three-node proposition (M1!) that some object (B1) is (i.e., has the property) red (RED).

cognitive agent) that communication can ever be "perfect". I, however, *can* understand my language of thought: directly, via syntactic understanding—"the mental structures applied in cognitive semantics *are* the meanings of the linguistic idioms; there is no further step of translating conceptual structure to something outside the mind" (as Gärdenfors 1993 puts it).

Recall that if Oscar does not understand Cassie, he cannot just reprogram her, or even read her mind (*pace* Carnap). Rather, he must repair the misunderstanding via language. Negotiation makes language understanding "self-correcting"; i.e., errors in language understanding can be corrected by further use of language. Negotiation does not *guarantee* mutual understanding, but it makes it possible—indeed, likely—and it makes residual misunderstands of marginal relevance.

The latter happens in two ways. In what follows, I will be talking about Cassie and Oscar simultaneously as computational models of (human) minds and as if they were computational cognitive agents whose minds are implemented in the SNePS semantic-network knowledge-representation and reasoning system. In a SNePS network, nodes represent either concepts or propositions, and labeled, directed arcs structure the nodes into grammatically well-formed, complex ("molecular") concepts and propositions (see, e.g., Shapiro & Rapaport 1987, 1995).

First, suppose for the sake of argument that Cassie's and Oscar's mental networks of concepts and beliefs differ only minimally. Suppose, for example, that Cassie believes that some object is red, and suppose that Oscar doesn't. In SNePS, Cassie's belief would be represented by an OBJECT-PROPERTY proposition, as in Figure 3. There would be three nodes: one for the object (B1), one for the property (RED), and one (M1!) for the proposition that B1 is (an object that has the property) red.[14] Note that Cassie's and Oscar's mental semantic networks need

[14] A SNePS node whose identifier is marked with '!' is said to be "asserted", i.e., believed by the cognitive agent whose mind is (being represented by) the SNePS network; cf. Rapaport et al.

only differ in *one* node (viz., M1); they might both share all other beliefs about red things and about B1. (This is, admittedly, implausible for all but the case of toy computational cognitive agents such as Cassie and Oscar, but it will serve to make my point.) Then, if Cassie tells Oscar something about some other object, Oscar will not fully appreciate all the connotations of Cassie's claim, because her claim will be linked to the 3-node proposition that Oscar lacks. (It might be linked directly, if her claim is about B1 or some other red thing, or it might be linked indirectly, if there is a long path connecting a node in her claim to either B1 or RED.) But in such a case of minimal belief difference, what Oscar misses will likely be of only marginal concern and, hence, irrelevant.

Second, suppose, again for the sake of argument, that Cassie's and Oscar's mental networks are structurally the same but that they differ only in some of the nodes representing the words that express their concepts. (In SNePS, these are the nodes at the heads of lex arcs, which are arcs that relate (atomic) concepts to terms expressing them. For more details on the semantics of lex arcs, see Shapiro & Rapaport 1987.) We might then have the following situation:

> Nicolaas de Bruijn once told me roughly the following anecdote: Some chemists were talking about a certain molecular structure, expressing difficulty in understanding it. De Bruijn, overhearing them, thought they were talking about mathematical lattice theory, since everything they said could be—and was—interpreted by him as being about the mathematical, rather than the chemical, domain. He told them the solution of their problem in terms of lattice theory. They, of course, understood it in terms of chemistry. (Rapaport 1995, §2.5.1.)

So, suppose that Cassie is discussing mathematical lattice theory but that Oscar is discussing chemistry, or that Cassie and Oscar are both apparently discussing a battle between two opposing armies, but Cassie is talking about chess while Oscar is discussing a battle in the Civil War.[15] As long as the two agents' interpretations of each other's utterances are isomorphic, neither will be able to determine that they are not talking about the same thing. Oscar, for instance, might not have the "intended interpretation" of Cassie's utterances; but this will make no practical difference:

> Jan and Edwige never understood each other, yet they always agreed. Each interpreted the other's words in his own way, and they lived in perfect harmony, the perfect solidarity of perfect mutual misunderstanding. (Kundera 1997, §3.1.; Rapaport 1998, §3.

[15] I think this example is due to John Haugeland, but I cannot track it down.

1978: 227. Quine's theories of radical translation (1960) and ontological relativity (1969) hover strongly in the background here.)

Lynne Rudder Baker calls these 'crazy interpretations' (personal communication, 21 April 1989). *Perhaps* Cassie and Oscar could calibrate their interpretations by reference to the real world. But I argue that this is not accessible to them (see §8 below, and Rapaport 2000b); any apparent such access is all internal. Hence, I cannot rule out crazy interpretations. But the need for successful communication makes such crazy interpretations irrelevant. Cassie and Oscar exist in a social environment, which constrains (or helps to constrain) the possible interpretations, even though it cannot rule out such "inverted spectrum" cases as where Cassie might be talking about a mathematical lattice and Oscar might understand her to be talking about the chemical structure of some molecule. Because Cassie and Oscar share a social environment, these differences will be irrelevant insofar as they have no pragmatic implications.[16] And to the extent that they *do* have pragmatic implications, they will make the need for negotiation evident.

How, then, does negotiation work? How are mistakes detected and corrected? By a continual process of: (1) hypothesis formation (the basic process of understanding); (2) hypothesis testing, some of which takes the form of further communication between the interlocutors (usually in the form of questions: "By 'X', did you mean Y?"); (3) belief revision based on the results of the tests; and (4) subsequent learning. The more we communicate with each other, the more we learn. We can ask questions and match the actual answer with our hypothesized one, or we can make trial statements and match our interlocutor's response with one we expect. If the question is not answered as we expect, or if the reply is surprising, we revise our beliefs. By successive approximation, we can asymptotically approach mutual comprehension (cf. Rapaport 1976: 178–180, Rapaport 1985/1986: 84–85).

Negotiation, moreover, is computable and hence can be part of a computational theory of mind and language. In addition to Sidner's work (cited above), Graeme Hirst and his colleagues have developed a computational theory of how humans can negotiate meanings in the service of correcting conversations that risk going astray (Hirst et al. 1994, McRoy & Hirst 1995, Hirst 2002). And

[16] And therefore they will be logically equivalent, or so Edwin Martin argued in unpublished lectures at Indiana University in Spring 1973. They can have pragmatic implications as the result of "joint sensory and manipulative acts", as my colleague Stuart C. Shapiro has pointed out to me. Because of the first-person point of view, however, the mental representations of these acts are what really matter.

Anthony Maida has discussed relevant issues in agent communication, focusing on the detection and correction of "existential misconceptions" such as "referent misidentification" (Maida 1990, 1991, 1992; Maida & Tang 1994, 1996, 1997). There is also a large and growing literature on "agent communication languages" for systems of multiple computational agents, many of which might be equipped with different ontologies (cf. Chaib-Draa & Dignum 2002, and the other papers in their special issue of *Computational Intelligence* devoted to the topic). However, whereas Hirst, Maida, Sidner and their colleagues are primarily interested in human-human communication or human-computer communication, the agent-communication-language community is primarily interested in computer-computer interactions; these are more tractable, because the human supervisors of such systems have control over the basic assumptions of their systems. One possible exception is Reed et al. 2002, which discusses the importance and nature of negotiation among "autonomous agents" communicating in a formal agent communication language, but it is not immediately clear how their theory applies to cognitive agents. The question of how "social" (i.e., common) meaning "emerges" from individual or first-person-point-of-view meaning is also of relevance, but beyond our immediate scope (cf. Putnam 1975 and, especially, Gärdenfors 1993).[17]

8 Bruner's Theory of Negotiation and Language Acquisition.

Jerome Bruner's studies of language acquisition (1983) shed light on communication and negotiation. According to Bruner, children *interpret* and *negotiate* during acquisition of their first language:

> The negotiation [between adult and child] has to do, probably, least with syntax, somewhat more with the semantic scope of the child's lexicon, and a very great deal with helping make intentions clear and making their expression fit the conditions and requirements of the "speech community", i.e., the culture. ... The development of language ... involves two people negotiating. ... If there is a Language Acquisition Device [LAD], the input to it is not a shower of spoken language but a highly interactive affair shaped ... by some sort of an adult Language Acquisition Support System [LASS]. (Bruner 1983: 38–39.)

[17] For a useful bibliography on negotiation in language, see Oeller 1998.

In a passage that is virtually a summary of much that I have been urging, Bruner sets out an example of language acquisition by the child:

> . . . reference can *vary* in precision from a rather wooly vagueness to a proper singular, definite referring expression. Indeed, two parties to a conversation may refer to the "same" topic with widely different degrees of precision. The "electricity" that a physicist mother has in mind will not be the "same" as what her child comprehends when she warns him about getting a shock. Still the two may carry on about "electricity" in spite of this indefiniteness. Their conversational negotiation may even *increase* her child's definiteness. Truth is not all that is involved in such causal chains. The child's conception of electricity may be vacuous or even wrong, yet there is a joint referent that not only exists in such asymmetric conversations, but that can be developed both for its truth value and its definiteness. (Bruner 1983: 67–68.)

There is much to applaud in this passage, but also much to take issue with and make more precise. For example, by 'reference', Bruner must mean the *act* of referring, for reference as understood, say, in the Fregean way is an all-or-nothing affair: A word either refers or it doesn't. Reference cannot "vary in precision", but *acts* of referring *could*: Speakers can be more or less careful, more or less sloppy, more or less detailed in their use of words. Further along, the fact that Bruner chose to use scare quotes when stating that the two speakers "may refer to the *'same'* topic" suggests that, indeed, "the" topic is *not* the "same", or else that the referring expressions used are associated with "widely different" concepts. So, again, he is not talking about the external, Fregean referent of the word.

What the physicist mother and her child do have that are "widely different" are their internal concepts associated with their common word 'electricity'. Indeed, the physicist will have a vast, complex network of concepts (even vaster than the ordinary adult), whereas initially the child will have none (it will be "vacuous"), or, at best, the child will have a concept of something—he or she knows not what—called 'electricity'. (This otherwise vacuous atomic concept expressed in English by 'electricity' is what is represented in SNePS by the node at the *tail* of a `lex` arc.) What *is* the "same" is the *referring term*, part of the public communication language; the associated (mental) concepts are *different*.

There is no place (so far) in Bruner's description for the external referent—electricity—itself. So, when mother and child "carry on about 'electricity' ", are they both talking about electricity itself? No, or not necessarily: "Truth is not all that is involved in such causal chains." Rather, they are talking "about" the word (i.e., the *t* of Fig. 2). Here, one must be careful not to confuse use with mention:

The only thing in common is the word. There are *two*, *distinct* meanings for the word: the physicist-mother's meaning and the child's meaning (i.e., the C_A and C_B of Fig. 2). The goal—in the long term—is for the child's meaning to be as much like the mother's as makes no difference. (In the case at hand, this may in fact be too much to ask, since most parents are not physicists. So the goal need only be for the child's meaning to be as much like an ordinary adult's meaning as makes no difference.) As the "conversational negotiation" continues, the child's concept will become more detailed, approaching that of the mother.

"Truth is not *all* that is involved", but is it involved at all? Bruner does say, at the end, that "there is a joint referent", but what is that joint referent? One might expect it to be electricity itself (i.e., the C of Fig. 2). But Bruner gets a bit murky here. He says that the joint referent "can be developed ... for its truth value and its definiteness". If so, then it cannot be electricity itself, because electricity itself has no "truth value"; it would be a category mistake to say so, for only sentences have truth values. However, the *term* 'electricity' does have the next best thing: a Fregean *referent*. This is consistent with the view developed in the previous paragraph.

Another argument against electricity itself being Bruner's "joint referent" is that electricity is neither "definite" nor "indefinite". But our theories or *concepts of* electricity *can* be. So, could the joint referent be the *common concept* of electricity that the mother hopes will be established? If so, why should Bruner—or we—think there must be such a thing? What, indeed, would it be? What is in common is indeed only the *word*; there are two *different* mental concepts associated with it, and the *child's* "can be developed". There is no need for a common concept or a common external object (cf. my discussion of Potts 1973 in Rapaport 2002). Rather, the picture we get from Bruner's description is this: There is something "joint", namely, the *word* 'electricity'. And there is a referent—in fact, two referents: Each person uses the *same word* to "refer" to his or her *own concept*; by negotiation, the concepts come into alignment.

There *is*, in a sense, *something* shared. Bruner says that "the means [for referring, or perhaps for the intent to refer] comprise the set of procedures by which two people establish 'jointness' in their attention" (Bruner 1983: 68). In what sense is their attention "joint"? Perhaps in the sense that what *I* am thinking of is what *you* are thinking of, though the 'is' here need not be the "is" of identity—it is more likely the "is" of equivalence or correspondence, as in node `M12` of Figure 4. That figure shows a SNePS representation of a situation, similar to the one Bruner describes, in which (1) I believe that something called 'electricity' shocks (node `M4!`); (2) I believe that *you* (my interlocutor) believe that something called

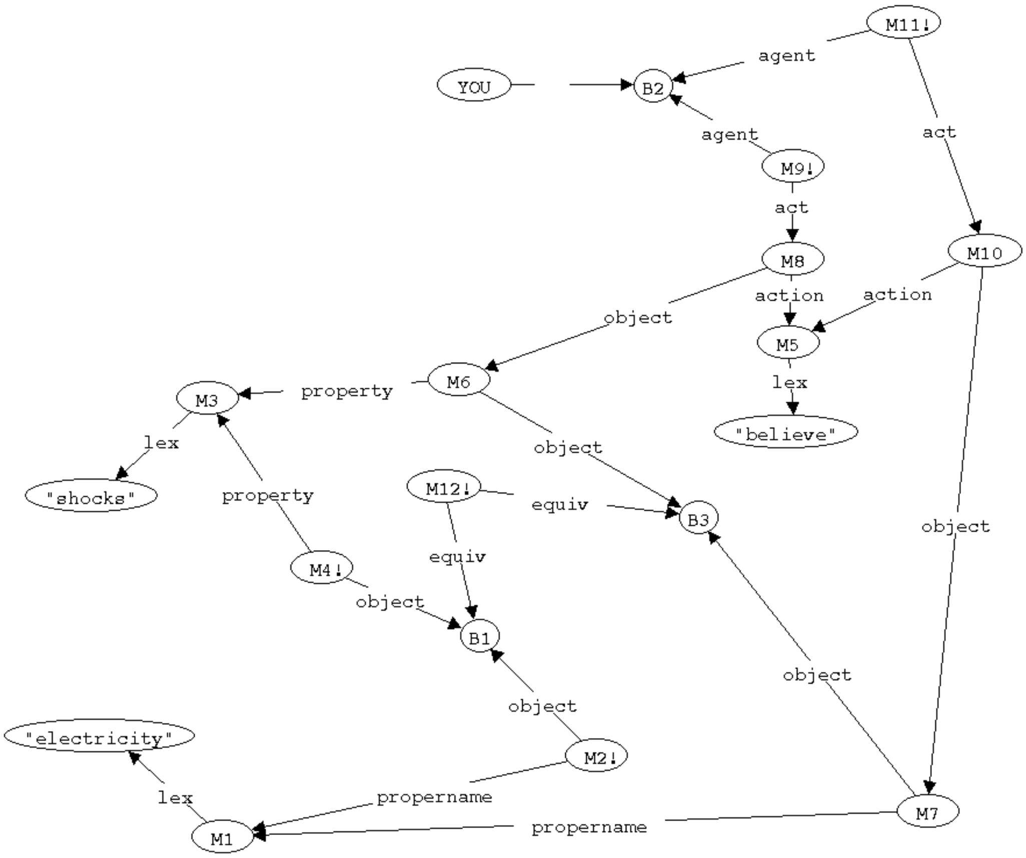


Figure 4: `M2!` = (I believe that) `B1` is called 'electricity';

`M4!` = (I believe that) `B1` shocks (i.e., I believe that there is something (viz., `B1`) called 'electricity' and that it shocks).

`M7` = `B3` is called 'electricity';

`M6` = `B3` shocks;

`M9!` & `M11!` = (I believe that) `B2` believes `M6` and `M7` (i.e., I believe that *you* (viz., `B2`) believe that there is something (viz., `B3`) called 'electricity' and that it shocks).

`M12!` = (I believe that) `B3` = `B1` (i.e., I believe that what you call 'electricity' is what I call 'electricity').

(The "you"-pointer mechanism is based on the I-pointer of Rapaport, Shapiro, & Wiebe 1997.)

'electricity' shocks (node `M9!`); and (3) I believe that what you call 'electricity' is what I call 'electricity' (node `M12!`). That is, what I am thinking of is what I believe that you are thinking of. What is "shared" is all in *one* mind (either the child's or the mother's)—shared by virtue of one of them having *both* his or her *own* concept *as well as* his or her own *representation of* the *other's* concept, plus some way of comparing them.

It is tempting to say that your intensional concept of something called 'electricity' corresponds to the same external object in the real world that *my* intensional concept corresponds to, but this temptation must be resisted, since that is really only sayable from a third-person perspective of someone with simultaneous access to the contents of our minds *and* the external world (cf. Rapaport 2000b), which is an impossible "view from nowhere" (to borrow Nagel's (1986) phrase). Such a comparison cannot be made "across" two distinct minds. A concept in one mind may indeed be similar to (or identical to, though numerically distinct from) a concept in another mind. But who has the warrant to make such a claim, except, perhaps, God, who might have simultaneous access to the contents of both minds? The only way for us poor mortals to make such a claim is within our own minds. I *can* compare two concepts that are both in *my* mind. If one of them is *my* representation of *your* concept, then (and only then) I can compare mine with "yours" (more accurately, with my representation of yours). This is what the theory of syntactic semantics calls the "internal (or 'narrow'), first-person point of view".

So, there is no need either for a joint *external* referent or for a joint *internal* referent. There is only need for sufficient similarity of structure of each conversant's internal networks for conversation to continue successfully:

> Achieving the *goal* of referring has little to do with agreement about a singular definite referent. It is enough that the parties to a referential exchange know that they share enough overlap in their focal attention to make it worthwhile continuing .... When the physicist mother tells her four-year-old that he has just been shocked by "electricity", she does not and need not assume that he has either the same extension or intension of the concept as she does. Nor need she care, if the conversation can only continue.
>
> The problem of how reference develops can, accordingly, be restated as the problem of how people manage and direct each other's attention by linguistic means. (Bruner 1983: 68.)

This notion of a speaker directing the hearer's attention is important for the issue of whether non-humans (including both computers as well as other animals) can understand and use language (cf. Terrace 1985), but it is beyond the scope of the present essay (see, however, Rapaport 1996, §9.4). For now, note that the picture we have, both from our own theory and from Bruner's, is this:

1. A cognitive agent A, communicating with another cognitive agent B, refers to (communicates a reference to?) some object C (or some mental concept C_A) via term t iff A directs B's attention to think about (for example, to find or else build in B's mental semantic network) a C-concept expressed by t (for example, a concept at the tail of a `lex` arc to t). Call it C_B. (C_B might represent C, or it might represent C_A.)

2. Moreover, if the communication is successful, C_A and A's representation of C_B will be more or less equivalent, and C_B and B's representation of C_A will be more or less equivalent, where the equivalence becomes "more" rather than "less" by negotiation.

I'll conclude this discussion of Bruner with one more quotation (but see Rapaport 1996, §9.6, for further discussion):

> John Lyons ... entitles an essay "Deixis as the Source of Reference." ... I think an equally strong case [can] ... be made ... that discourse and dialogue are also the sources of reference. If they were not, each speaker would be locked in a web of isolated referential triangles of his [sic] own making—if indeed he could construct such a web on his own. (Bruner 1983: 88.)[18]

That is, *negotiation* is a source of reference. More precisely, one way to get out of the "web of isolated referential triangles"[19]—to "ground" one's symbols—is by means of dialogue. Note, however, that, even on Bruner's own view, dialogue does not really get us "out" of our internal network, since all it can do is set up correspondences in each speaker's mind between objects in two belief spaces: the speaker's and the speaker's representation of the hearer's (together with correspondences with internal representations of external objects).

[18]For work on deixis related to syntactic semantics, cf. Duchan et al. 1995.

[19]Circles, however, seem to me a preferable geometric metaphor.

9 Understanding and Generating Language.

What happens in communication? When I speak—when I *generate* an utterance—I generate expressions "that are pertinent to the [neurophysiological] stimulus and are usable to narrate the primary [neurophysiological] display when inserted in appropriate grammatical structures" (Damasio 1989: 25). For example, I conceive or perceive an object, which causes neuronal activity representing its features and structure. This activity is linked to other neuronal structures that "generate names" (Damasio 1989: 25) that, in turn, allow me to communicate two things to you: (1) *that* I am thinking of an object and (2) *what* it is.

But "I cannot be sure that the meaning of a word I say is the same for the person to whom I direct it" (Vauclair 1990: 321). And there is an equal but opposite uncertainty on the part of the hearer. As my colleague Jorge J.E. Gracia has expressed it,

> We do not perceive ideas; what we perceive are certain phenomena that *suggest to us* certain ideas. If I ask you, for example, "Do you approve of what the President did?" and you frown in return, I conclude that you do not. But it is altogether possible that you do in fact approve . . . , although you . . . mislead me by making the frown. *My conclusion* that you do not, then, can be taken only as an *interpretation* of what you are thinking based on certain empirical evidence that is only *indirectly related to what you think*. (Gracia 1990: 495; my emphases.)

I can't have direct access to your thoughts, only to your language acts and gestures. My interpretation is not of what you are thinking, but of your language and gestures. It is, as noted above (§5), a defeasible conclusion.

Possibly, however, symbols don't "*convey* meaning". Here, I am using 'convey' in the sense found in a *Calvin & Hobbes* cartoon (22 February 1990) in which Calvin (a small boy) has just built an abstract "snowman" with a large hole in it. The following dialogue ensues between Calvin and his imaginary tiger friend, Hobbes:

> Hobbes: How's your snow art progressing?
> Calvin: I've moved into abstraction!
> Hobbes (looking with puzzlement at the "snowman"): Ah.
> Calvin: This piece is about the inadequacy of traditional imagery and symbols to convey meaning in today's world. By abandoning representationalism, I'm free to express myself with pure form. Specific interpretation gives way to a more visceral response.

Hobbes: I notice your oeuvre is monochromatic.
Calvin: Well c'mon, it's just snow.

Instead of the speaker's symbols (expressions) *conveying* meaning—and instead of the hearer having to *interpret* the expressions—they *elicit* meaning in the hearer's mind (i.e., they act as stimuli to "activate" concepts; they produce a "visceral response"). With luck and negotiation, the ideas elicited or activated in the hearer's mind are constrained by the context and the dialogue to be structurally similar to the speaker's ideas expressed by the speaker's symbols.

Again, it is tempting to say that, because of structural similarity, the hearer's ideas correspond to the same things that the speaker's correspond to. But, again, the temptation to use such external, third-person modes of expression must be resisted. In addition to the reasons discussed in the previous section, there might not *be* anything in the external world for our ideas to correspond to: Figure 2's *C* might not exist. Yet our language behavior is no different in the case when *C* does exist than in the case when it does not (cf. Rapaport 1981).

10 Winston's Problem for Knowledge Representation.

If a lion could talk, we could not understand him. (Wittgenstein 1958: 223.)

Will negotiation always work? If a computer could talk, would we understand each other? Cognitive agents with different (types of) bodies might have different concepts; we would literally be thinking different things. (As in a *New Yorker* cartoon showing a man waiting for a green light so that he can cross the street; the man is thinking "Work. Eat. Sleep. Work. Eat. Sleep.", while the traffic light is thinking "Green. Yellow. Red. Green. Yellow. Red." Or as in a Gary Larson *Far Side* cartoon, "How birds see the world", showing a bird's-eye view of a man, a woman, and a dog, each with a target superimposed on them.) The concepts of such cognitive agents would be thoughts nonetheless, but such differences might make mutual comprehension impossible. I will call this 'Winston's Problem', in honor of a formulation of it by Patrick Henry Winston in his early work on machine learning (though arguably the honor should go to Wittgenstein):

> Simulation of human intelligence is not a primary goal of this work. Yet for the most part I have designed programs that see the world in terms conforming to human usage and taste. These programs produce descriptions that use notions such as left-of, on-top-of, behind, big, and part-of.
>
> There are several reasons for this. One is that if a machine is to learn from a human teacher, then it is reasonable that the machine should understand and use the same relations that the human does. Otherwise there would be the sort of difference in point of view that prevents inexperienced adult teachers from interacting smoothly with small children.
>
> Moreover, if the machine is to understand its environment for any reason, then understanding it in the same terms humans do helps us to understand and improve the machine's operation. Little is known about how human intelligence works, but it would be foolish to ignore conjectures about human methods and abilities if those things can help machines. Much has already been learned from programs that use what seem like human methods. There are already programs that prove mathematical theorems, play good chess, work analogy problems, understand restricted forms of English, and more. Yet, in contrast, little knowledge about intelligence has come from perceptron work and other approaches to intelligence that do not exploit the planning and hierarchical organization that seems characteristic of human thought.
>
> Another reason for designing programs that describe scenes in human terms is that human judgment then serves as a standard. There will be no contentment with machines that only do as well as humans. But until machines become better than humans at seeing, doing as well is a reasonable goal, and comparing the performance of the machine with that of the human is a convenient way to measure success. (Winston 1975/1985: 143; cf. Kirsh 1991: 22–24, Maida 1985).

Winston's Problem concerns what might happen if the knowledge-representation language (i.e., language of thought) of a computer system that can learn concepts differs significantly from that of humans. According to Winston (and Wittgenstein), what would happen is that the two systems—computer and human (or lion and human)—would not be able to understand each other. How serious is this problem? Quite serious, according to Joseph Weizenbaum, who observed that the intelligence of computers "must always be an intelligence *alien* to genuine human problems and concerns" (1976: 213).

There are reasons to be optimistic, however. Winston's Problem comes in different degrees:

1. Consider two computational cognitive agents, Cassie and Oscar, who share both a public communication language (say, English) and a language of thought. For concreteness, suppose their language of thought is the SNePS/Cassie knowledge-representation language (as described in Shapiro & Rapaport 1987). Winston's Problem would arise here only to the extent that it arises for any of us in everyday life: I, as a male, can never experience pregnancy; so, my understanding of 'pregnant' is qualitatively different from that of a female (certainly from that of a female who has been pregnant). Yet I use the word, am not misunderstood when I use it, and can understand (within recognized limits) a woman's use of it.[20] Insofar as our *experiences* differ—insofar as we have different background or "world" knowledge—then to that extent will we mutually misunderstand each other. As we have seen, though, the more we communicate—and thereby negotiate—the more we will come to understand each other.

2. If Cassie and Oscar share only a language of thought, but *not* a public communication language, then there is an extra layer of difficulty due to the difficulties of translation. Still, with enough work, dialogue, and explanatory glosses, this can be overcome (cf. Jennings 1985, Rapaport 1988).

3. In either of the above cases, things would be made worse if Cassie's and Oscar's "conceptual schemes" differ. By this, I don't mean that their languages of thought differ, but that their "world knowledge" is so different that even common experiences would be differently interpreted—and radically so:

> The falling of a maple leaf is a sign of autumn ... because *we* have established a connection between them on the basis of certain observations and, therefore, use the phenomena in question to indicate something of interest to us. A different culture ... might see the falling of a maple leaf ... as [a] sign of other events or even as [an] indication of the divine will to punish and reward them. (Gracia 1990: 502; my italics.)

And anthropologists tell us that where Western physicians see viruses and bacteria, other cultures, such as the Kulina of Brazil, see *dori*—a substance "that permeates the flesh of shamans, giving them the ability to cure as well as to injure others"—injected into the body of a victim by a shaman (Pollock 1996: 329). This is not unlike the situation discussed earlier where there is a single computer program with two distinct input-output encodings, so that one computer is taken to be discussing chess while the other is taken to be discussing a Civil War battle

[20]The example is Shapiro's and is discussed further in Rapaport 1988: 116, 126n20.

(or one is taken to be discussing chemistry, the other, mathematical lattice theory). Here, Winston's Problem begins to get a bit more serious. Nonetheless, it appears that we can *understand* the other's point of view, even if we disagree with it.

4. Winston's Problem becomes more threatening, of course, when the languages of thought differ. Even here, there are degrees of difference. For instance, Cassie and Oscar might both have SNePS languages of thought, but Cassie's might use the case frames (i.e., sets of arc labels) that Shapiro and I advocate (Shapiro & Rapaport 1987) whereas Oscar might use those advocated by Richard Wyatt (1990, 1993). Here we have an empirically testable hypothesis that, say, one of the languages of thought would be "better" than the other in the sense of enabling the cognitive agent whose language of thought it is to understand finer discriminations. "All" we would have to do to test it is implement Cassie using the Shapiro-&-Rapaport case frames, implement Oscar using the Wyatt case frames, and then let them converse with each other. (I put 'all' in scare quotes, because, obviously, we would also have to do lots of other things, such as implement understanding and generation grammars, give Cassie and Oscar background knowledge, and devise appropriate test dialogues.) Conceivably, one of the languages of thought might be so (relatively) impoverished that its "user" would simply not be able to understand or express some distinction that the other could.

5. Another level of difficulty—equally empirically testable—would arise if the two languages of thought were distinct members of the same general *kind* of knowledge-representation language. For instance, we might run our experiment with both Cassie and Oscar implemented in different symbolic, intensional, knowledge-representation and reasoning systems, say, Cassie in (some version of) SNePS and Oscar in (some version of) KL-ONE (Brachman & Schmolze 1985, Woods & Schmolze 1992).

6. The *potential* for more serious inability to communicate occurs when one of the computational cognitive agents has a *connectionist* language of thought while the other has a "classical" symbolic one. (A version of Winston's Problem arises in those connectionist models in which it is not at all clear what, if anything, the final weights on the connections "mean" in terms of the task that the system has learned.) This, I take it, is the situation Winston had in mind when he referred to work on perceptrons, though he wrote before connectionism was as well-investigated as it is now. There would indeed be a problem if Cassie, whose language of thought was symbolic, tried to "read the mind" of an Oscar whose language of thought was connectionist. But as long as they spoke a common, public communication language, negotiation via dialogue might overcome any residual problems. This, too, is testable. Indeed, for all we know, we are testing just such hypotheses as this and the ones proposed in points 4 and 5, above, every day when we speak!

7. The worst case would be what I will call the "Black Cloud case": Consider as simple a term as 'in'. George Lakoff argues that it is human-bodily centered, based on our knowledge of the inside and outside of our bodies (1987: 271–273). But consider a cognitive agent whose body is a "black cloud" in the style of Fred Hoyle's (1957) novel. (Such a science-fiction case is necessary here, since the whole point is that if *we* can't imagine how 'in' could be non-objective, then, to imagine it, we need a non-human example.) The Black Cloud, not having an inside, might not have a concept of "in". How would such a cognitive agent describe a pea *in* a cup? Topologically, the pea is *on* the cup. So, perhaps, "on" is an objective concept. No matter. I would venture that such remaining objective relations are too few to describe the world. Another example: What about 'inside', as in a pea *inside* a closed box? Perhaps one has no concept of "inside" the box, but the box makes noises if shaken, and, if opened, one sees that now there is a pea *on* the box (in the topological sense of 'on'). Note how hard it would be for the Black Cloud to translate human language (or, at least, English). Here there is no common *kind* of language of thought, no common conceptual scheme, no common public communication language. This would appear to be a case for despair, though some are optimistic (e.g., Hoyle himself, and Sagan 1980: 287–289). The optimism, it should be noted, comes from the hope that there is *enough* of a common basis to get negotiational dialogue off to a start. (Some, such as McAllister 2001, think that some innate concepts are needed in addition to negotiation.)

How much of a common basis is needed? Are computers (or computational cognitive agents) so "alien" to us (as Weizenbaum says) that there is little hope for a sufficiently large common basis? Or, as perhaps Searle (1980) and others would have it, so alien that there is little hope that computational theories of natural-language understanding and generation will ever reach the level of real understanding? Perhaps Lakoff (1987) is correct that, for there to be any hope of avoiding Winston's Problem, a robot will have to have a human-like body, i.e., be an android. But how would we know what an android's concepts are?[21] For that matter, how do I know what *your* concepts are? What consideration of these cases suggests is that Winston's Problem can be overcome as long as there is a public communication language and as long as we are able and willing to negotiate in it.

[21] As Dick 1968 asked, "Do androids dream of electric sheep?".

11 Conclusion.

> *A book is a way to hold the mind of another in your hands. You can have a dialogue with Plato. . . . Books. How you reach across time and space to be with another's mind.* (Advertisement for Doubleday Book Shops, *The New Yorker* 67 (18 November 1991) 111.)

But minds are abstract (brains and computers are physical). To be able to be in causal communication with a mind, its ideas need to be implemented—to be expressed in a syntactic medium that can subsequently be (re-)interpreted in, or by, another mind.

When we communicate, we attempt to convey our internal meanings to an audience (a hearer or reader) by means of a public communication language: "A book is a way to hold the mind of another in your hands." Paradoxically, this attempt almost always both fails and nearly succeeds. Near misunderstandings can be ignored. Larger ones can be minimized through negotiation—allowing us to approach, though perhaps never reach, complete mutual comprehension.

Acknowledgments

I am grateful to my colleagues Stuart C. Shapiro, J.P. Koenig, and the other members of the SNePS Research Group for their comments on an earlier draft. Preparation of this essay was supported by grant REC-0106338 from the National Science Foundation.

References

Alvarez, Celso (3 May 1990), article 6387 on the `sci.lang` electronic bulletin board.

Barker, Clive (1987), *Weaveworld* (New York: Poseidon).

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

Bruner, Jerome (1983), *Child's Talk: Learning to Use Language* (New York: W. W. Norton).

Campbell, Alistair E. (1999), *Ontological Mediation: Finding Translations across Dialects by Asking Questions* Ph.D. dissertation (Buffalo: SUNY Buffalo Department of Computer Science).

Campbell, Alistair E., & Shapiro, Stuart C. (1998), "Algorithms for Ontological Mediation", in S. Harabagiu (ed.), *Usage of WordNet in Natural Language Processing Systems: Proceedings of the [COLING-ACL] Workshop*: 102–107.

Carnap, Rudolf (1956), *Meaning and Necessity: A Study in Semantics and Modal Logic, 2nd edition* (Chicago: University of Chicago Press).

Chaib-Draa, B., & Dignum, F. (2002), "Trends in Agent Communication Language", *Computational Intelligence* 18(2): 89–101.

Damasio, Antonio R. (1989), "Concepts in the Brain", in "Forum: What is a Concept?", *Mind and Language* 4: 24–27.

Dick, Philip K. (1968), *Do Androids Dream of Electric Sheep?* (Garden City, NY: Doubleday).

Duchan, Judith Felson; Bruder, Gail A.; & Hewitt, Lynne E. (eds.) (1995), *Deixis in Narrative: A Cognitive Science Perspective* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Durfee, Edmund H. (1992), "What Your Computer Really Needs to Know, You Learned in Kindergarten", *Proceedings of the 10th National Conference on Artificial Intelligence (AAAI-92; San Jose, CA)* (Menlo Park CA: AAAI Press/MIT Press): 858–864.

Ehrlich, Karen (1995), "Automatic Vocabulary Expansion through Narrative Context", *Technical Report 95-09* (Buffalo: SUNY Buffalo Department of Computer Science).

Fodor, Jerry, & Lepore, Ernest (1991), "Why Meaning (Probably) Isn't Conceptual Role," *Mind and Language* 6: 328–343.

Fodor, Jerry, & Lepore, Ernest (1992), *Holism: A Shopper's Guide* (Cambridge, MA: Basil Blackwell).

Gärdenfors, Peter (1993), "The Emergence of Meaning", *Linguistics and Philosophy* 16: 285–309.

Gärdenfors, Peter (1997), "Meanings as Conceptual Structures", in M. Carrier & P. Machamer (eds.), *Mindscapes: Philosophy, Science, and the Mind* (Pittsburgh: University of Pittsburgh Press): 61–86.

Gärdenfors, Peter (1999a) "Does Semantics Need Reality?", in Alexander Riegler, Markus Peschl, & Astrid von Stein (eds.), *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?* (New York: Kluwer Academic/Plenum Publishers): 209–217.

Gärdenfors, Peter (1999b), "Some Tenets of Cognitive Semantics", in Jens Allwood & Peter Gärdenfors (eds.), *Cognitive Semantics: Meaning and Cognition* (Amsterdam: John Benjamins): 19–36.

Geeraerts, Dirk (1997), *Diachronic Prototype Semantics: A Contribution to Historical Linguistics* (Oxford: Clarendon Press).

Győri, Gábor (2002), "Semantic Change and Cognition", *Cognitive Linguistics* 13(2): 123–166.

Gracia, Jorge J. E. (1990), "Texts and Their Interpretation", *Review of Metaphysics* 43: 495–

542.

Harris, Albert J., & Sipay, Edward R. (1990), *How to Increase Reading Ability: A Guide to Developmental and Remedial Methods; ninth edition* (New York: Longman).

Hirst, Graeme (2002), "Negotiation, Compromise, and Collaboration in Interpersonal and Human-Computer Conversations", *Proceedings, Workshop on Meaning Negotiation, 18th National Conference on Artificial Intelligence (AAAI 2002, Edmonton)*: 1–4; online at [ftp://ftp.cs.toronto.edu/pub/gh/Hirst-MeanNeg-2002.pdf]

Hirst, Graeme; McRoy, Susan; Heeman, Peter; Edmonds, Philip; & Horton, Diane (1994), "Repairing Conversational Misunderstandings and Non-Understandings", *Speech Communication* 15(3–4): 213–229.

Hoyle, Fred (1957), *The Black Cloud* (New York: Harper & Row).

Jennings, Richard C. (1985), "Translation, Interpretation and Understanding," paper read at the American Philosophical Association Eastern Division (Washington, DC); abstract, *Proceedings and Addresses of the American Philosophical Association* 59: 345–346.

Kearns, John (forthcoming 2003), "The Logic of Coherent Fiction", in T. Childers et al. (eds.), *The Logica Yearbook 2002* (Prague: Filosofia).

Kirsh, David (1991), "Foundations of AI: The Big Issues", *Artificial Intelligence* 47: 3–30.

Kundera, Milan (1978), *The Book of Laughter and Forgetting*, trans. by Michael Henry Heim (New York: Penguin Books).

Lakoff, George (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press).

Lenat, Douglas B. (1995), "CYC, WordNet, and EDR: Critiques and Responses—Lenat on WordNet and EDR", *Communications of the ACM* 38.11 (November) 45–46.

Maida, Anthony S. (1985), "Selecting a Humanly Understandable Knowledge Representation for Reasoning about Knowledge", *International Journal of Man-Machine Studies* 22(2): 151–161.

Maida, Anthony S. (1990), "Dynamically Detecting Existential Misconceptions", *Technical Report CS-90-39* (University Park, PA: Pennsylvania State University Department of Computer Science).

Maida, Anthony S. (1991), "Maintaining Mental Models of Agents Who Have Existential Misconceptions", *Artificial Intelligence* 50(3): 331–383.

Maida, Anthony S. (1992), "Knowledge Representation Requirements for Description-Based Communication", in Bernhard Nebel, Charles Rich, & William Swartout (eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the 3rd International Conference (KR '92)* (San Mateo, CA: Morgan Kaufmann): 232–243.

Maida, Anthony S. (1996), "Referent Misidentification and Recovery among Communicating Agents", in Edmund Durfee (ed.), *Proceedings of the 2nd International Conference on Multiagent Systems (ICMAS-96)*: 196–203.

Maida, Anthony S., & Shapiro, Stuart C. (1982), “Intensional Concepts in Propositional Semantic Networks”, *Cognitive Science* 6: 291–330; reprinted in Ronald J. Brachman & Hector J. Levesque (eds.), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kaufmann, 1985): 169–189.

Maida, Anthony S., & Tang, Shaohua (1994), “Knowledge Management to Support Description-Based Communication for Autonomous Agents”, *Proceedings of the 7th Florida AI Research Symposium*: 184–188.

Maida, Anthony S., & Tang, Shaohua (1997), “Description-Based Communication for Autonomous Agents under Ideal Conditions”, *Journal of Experimental and Theoretical Artificial Intelligence* 9: 103–135.

McAllister, Destanie (2001), “Review of *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind* by Jerry Fodor”, *Cognitive Science Society Newsletter* 21(3) (September), [http://www.cognitivesciencesociety.org/newsletter/Sept01/fodrev.html]

McGilvray, James (1998), “Meanings Are Syntactically Individuated and Found in the Head”, *Mind and Language* 13: 225–280.

McRoy, Susan W., & Hirst, Graeme (1995), “The Repair of Speech Act Misunderstandings by Abductive Inference”, *Computational Linguistics* 21(4): 435–478.

Melzack, Ronald (1992), “Phantom Limbs”, *Scientific American* 266 (April): 120–126.

Nagel, Thomas (1986), *The View from Nowhere* (New York: Oxford University Press).

Neisser, Ulric (1976), *Cognition and Reality: Principles and Implications of Cognitive Psychology* (San Francisco: Freeman).

Oeller, Wilfried (1998, November 19), “Negotiation in Interaction”, *LINGUIST List* 9.1645 [http://www.linguistlist.org/issues/9/9-1645.html].

Pollock, Donald (1996), “Personhood and Illness among the Kulina”, *Medical Anthropology Quarterly* N.S.10(3): 319–341.

Posner, Roland (1992), “Origins and Development of Contemporary Syntactics,” *Languages of Design* 1: 37–50.

Potts, Timothy C. (1973), “Model Theory and Linguistics”, in Edward L. Keenan (ed.), *Formal Semantics of Natural Language* (Cambridge, UK: Cambridge University Press, 1975): 241–250.

Putnam, Hilary (1975), “The Meaning of ‘Meaning’,” reprinted in *Mind, Language and Reality* (Cambridge, UK: Cambridge University Press): 215–271.

Quillian, M. Ross (1967), “Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities”, *Behavioral Science* 12: 410–430; reprinted in Ronald J. Brachman & Hector J. Levesque (eds.), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kaufmann, 1985): 97–118.

Quillian, M. Ross (1968), “Semantic Memory,” in Marvin Minsky (ed.) *Semantic Information Processing* (Cambridge, MA: MIT Press): 227–270.

Quillian, M. Ross (1969), “The Teachable Language Comprehender: A Simulation Pro-

gram and Theory of Language", *Communications of the Association for Computing Machinery* 12(8): 459–476.
Quine, Willard Van Orman (1960), *Word and Object* (Cambridge, MA: MIT Press).
Quine, Willard Van Orman (1969), "Ontological Relativity," in W.V. Quine, *Ontological Relativity and Other Essays* (New York: Columbia University Press): 26–68.
Rapaport, William J. (1976), *Intentionality and the Structure of Existence*, Ph.D. dissertation (Bloomington: Indiana University Department of Philosophy).
Rapaport, William J. (1981), "How to Make the World Fit Our Language: An Essay in Meinongian Semantics", *Grazer Philosophische Studien* 14: 1–21.
Rapaport, William J. (1985/1986), "Non-Existent Objects and Epistemological Ontology", *Grazer Philosophische Studien* 25/26: 61–95; reprinted in Rudolf Haller (ed.), *Non-Existence and Predication* (Amsterdam: Rodopi, 1986).
Rapaport, William J. (1986), "Searle's Experiments with Thought", *Philosophy of Science* 53: 271–279; preprinted as *Technical Report 216* (Buffalo: SUNY Buffalo Department of Computer Science, 1984).
Rapaport, William J. (1988), "Syntactic Semantics: Foundations of Computational Natural-Language Understanding", in James H. Fetzer (ed.), *Aspects of Artificial Intelligence* (Dordrecht, Holland: Kluwer Academic Publishers): 81–131;
errata, [http://www.cse.buffalo.edu/∼rapaport/Papers/synsem.original.errata.pdf]).
Reprinted in Eric Dietrich (ed.), *Thinking Computers and Virtual Persons: Essays on the Intentionality of Machines* (San Diego: Academic Press, 1994): 225–273;
errata, [http://www.cse.buffalo.edu/∼rapaport/Papers/synsem.reprint.errata.pdf]).
Rapaport, William J. (1995), "Understanding Understanding: Syntactic Semantics and Computational Cognition", in James E. Tomberlin (ed.), *Philosophical Perspectives, Vol. 9: AI, Connectionism, and Philosophical Psychology*, (Atascadero, CA: Ridgeview): 49–88; reprinted in Josefa Toribio & Andy Clark (1998), *Artificial Intelligence and Cognitive Science: Conceptual Issues, Vol. 4: Language and Meaning in Cognitive Science: Cognitive Issues and Semantic Theory*, (Hamden, CT: Garland): 73–88.
Rapaport, William J. (1996), *Understanding Understanding: Semantics, Computation, and Cognition*, *Technical Report 96-26* (Buffalo: SUNY Buffalo Department of Computer Science);
on line at either: [ftp://ftp.cse.buffalo.edu/pub/tech-reports/96-26.ps]
or [http://www.cse.buffalo.edu/∼rapaport/Papers/book.ps]
Rapaport, William J. (1998), "How Minds Can Be Computational Systems", *Journal of Experimental and Theoretical Artificial Intelligence* 10: 403–419.
Rapaport, William J. (1999), "Implementation Is Semantic Interpretation", *The Monist* 82: 109–130; longer version preprinted as *Technical Report 97-15* (Buffalo: SUNY Buffalo Department of Computer Science) and *Technical Report 97-5* (Buffalo: SUNY Buffalo Center for Cognitive Science):
[http://www.cse.buffalo.edu/∼rapaport/Papers/implementation.pdf].

Rapaport, William J. (2000a), "Cognitive Science", in Anthony Ralston, Edwin D. Reilly, & David Hemmendinger (eds.), *Encyclopedia of Computer Science, 4th edition* (New York: Grove's Dictionaries): 227–233.

Rapaport, William J. (2000b), "How to Pass a Turing Test: Syntactic Semantics, Natural-Language Understanding, and First-Person Cognition", *Journal of Logic, Language, and Information* 9: 467–490.

Rapaport, William J. (2002), "Holism, Conceptual-Role Semantics, and Syntactic Semantics", *Minds and Machines* 12(1): 3–59.

Rapaport, William J., & Ehrlich, Karen (2000), "A Computational Theory of Vocabulary Acquisition", in Łucja M. Iwańska & Stuart C. Shapiro (eds.), *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language* (Menlo Park, CA/Cambridge, MA: AAAI Press/MIT Press): 347–375; errata, [http://www.cse.buffalo.edu/∼rapaport/Papers/krnlp.errata.pdf].

Rapaport, William J., & Kibby, Michael W. (2002), "Contextual Vocabulary Acquisition: A Computational Theory and Educational Curriculum", in Nagib Callaos, Ana Breda, and Ma. Yolanda Fernandez J. (eds.), *Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002; Orlando, FL)* (Orlando: International Institute of Informatics and Systemics), Vol. II: Concepts and Applications of Systemics, Cybernetics, and Informatics I, pp. 261–266.

Rapaport, William J.; Shapiro, Stuart C.; & Wiebe, Janyce M. (1997), "Quasi-Indexicals and Knowledge Reports", *Cognitive Science* 21: 63–107; reprinted in Francesco Orilia & William J. Rapaport (eds.), *Thought, Language, and Ontology: Essays in Memory of Hector-Neri Castañeda* (Dordrecht: Kluwer Academic Publishers, 1998): 235–294.

Reed, Chris; Norman, Timothy J.; & Jennings, Nicholas R. (2002), "Negotiating the Semantics of Agent Communication Languages", *Computational Intelligence* 18(2): 229–252.

Russell, Bertrand (1918), "The Philosophy of Logical Atomism", reprinted in Bertrand Russell, *Logic and Knowledge: it Essays 1901–1950*, R.C. Marsh (ed.), (New York: Capricorn, 1956): 177–281.

Sacks, Oliver W. (1989), *Seeing Voices: A Journey into the World of the Deaf* (Berkeley: University of California Press).

Sacks, Oliver (1990), "Seeing Voices", *Exploratorium Quarterly* Vol. 14, No. 2 (Summer), p. 3.

Sagan, Carl (1980), *Cosmos* (New York: Random House).

Saussure, Ferdinand de (1959), *Course in General Linguistics*, ed. by Charles Bally, Albert Sechehaye, & Albert Reidlinger, trans. by Wade Baskin (New York: Philosophical Library).

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

Shapiro, Stuart C. (1992), "Artificial Intelligence", in Stuart C. Shapiro (ed.), *Encyclopedia of Artificial Intelligence, second edition* (New York: John Wiley & Sons): 54–57.

Shapiro, Stuart C. (1993), "Belief Spaces as Sets of Propositions", *Journal of Experimental and Theoretical Artificial Intelligence* 5: 225–235.

Shapiro, Stuart C. & Ismail, Haythem O. (2001), "Symbol-Anchoring in Cassie", in Silvia Coradeschi & Alessandro Saffioti (eds.), *Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems: Papers from the 2001 AAAI Fall Symposium*, Technical Report FS-01-01 (Menlo Park, CA: AAAI Press): 2–8.

Shapiro, Stuart C., & Rapaport, William J. (1987), "SNePS Considered as a Fully Intensional Propositional Semantic Network", in Nick Cercone & Gordon McCalla (eds.), *The Knowledge Frontier: Essays in the Representation of Knowledge* (New York: Springer-Verlag): 262–315; shorter version appeared in *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86, Philadelphia)* (Los Altos, CA: Morgan Kaufmann): 278–283; a revised shorter version appears as "A Fully Intensional Propositional Semantic Network", in Leslie Burkholder (ed.), *Philosophy and the Computer* (Boulder, CO: Westview Press, 1992): 75–91.

Shapiro, Stuart C., & Rapaport, William J. (1991), "Models and Minds: Knowledge Representation for Natural-Language Competence," in Robert Cummins & John Pollock (eds.), *Philosophy and AI: Essays at the Interface* (Cambridge, MA: MIT Press): 215–259.

Sidner, Candace L. (1994), "An Artificial Discourse Language for Collaborative Negotiation", *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94, Seattle)* (Menlo Park, CA: AAAI Press/MIT Press): 814–819.

Simon, Herbert A. (1992), "The Computer as a Laboratory for Epistemology", in Leslie Burkholder (ed.), *Philosophy and the Computer* (Boulder, CO: Westview Press): 3–23.

Slocum, Jonathan (1985), "A Survey of machine Translation: Its History, Current Status, and Future Prospects", *Computational Linguistics* 11(1): 1–17.

Smith, Brian Cantwell (1987), "The Correspondence Continuum", *Report CSLI-87-71* (Stanford, CA: Center for the Study of Language and Information).

Steels, Luc (1998), "The Origins of Syntax in Visually Grounded Robotic Agents", *Artificial Intelligence* 103: 133–156.

Talmy, Leonard (2000), *Toward a Cognitive Semantics* (Cambridge, MA: MIT Press).

Terrace, Herbert S. (1985), "In the Beginning Was the 'Name'," *American Psychologist* 40: 1011–1028.

Traugott, Elizabeth C. (1999), "The Role of Pragmatics in Semantic Change", in Jef Verschueren (ed.), *Pragmatics in 1998: Selected Papers from the 6th International Pragmatics conference* (Antwerp: International Pragmatics Association), Vol. II, pp. 93–102.

Vauclair, Jacques (1990), "Primate Cognition: From Representation to Language", in

S. T. Parker & K. R. Gibson (eds.), *"Language" and Intellect in Monkeys and Apes* (Cambridge, UK: Cambridge University Press): 312–329.

Weizenbaum, Joseph (1976), *Computer Power and Human Reason: From Judgment to Calculation* (New York: W. H. Freeman).

Winston, Patrick Henry (1975), "Learning Structural Descriptions from Examples", in Patrick Henry Winston (ed.), *The Psychology of Computer Vision* (New York: McGraw-Hill): 157–209; reprinted in Ronald J. Brachman & Hector J. Levesque (eds.), *Readings in Knowledge Representation* (Los Altos, CA: Morgan Kaufmann, 1985): 141–168 (page references are to this reprint).

Wittgenstein, Ludwig (1958), *Philosophical Investigations, 3rd edition*, trans. by G.E.M. Anscombe (New York: Macmillan).

Wolfe, Maynard Frank (2000), *Rube Goldberg: Inventions* (New York: Simon & Schuster).

Woods, William A., & Schmolze, James G. (1992), "The KL-ONE Family", *Computers and Mathematics with Applications* 23: 133–177; reprinted in Fritz Lehmann (ed.), *Semantic Networks in Artificial Intelligence* (Oxford: Pergamon Press, 1992): 133–177.

Wyatt, Richard (1990), "Kinds of Opacity and Their Representations", in Deepak Kumar (ed.), *Current Trends in SNePS—Semantic Network Processing System*, Lecture Notes in Artificial Intelligence, No. 437 (Berlin: Springer-Verlag): 123–144.

Wyatt, Richard (1993), "Reference and Intensions", *Journal of Experimental and Theoretical Artificial Intelligence* 5: 263–271.

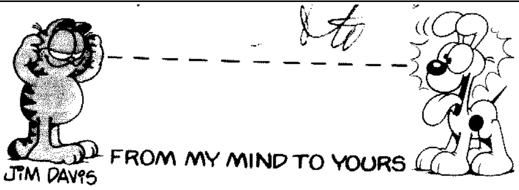


Figure 5: Information gets sent from the sender's (i.e., Garfield the cat's) mind to the recipient's (i.e., Odie the dog's) by being written on paper (that is, by being implemented in language). (From Post-It Note P-788, ©1978, United Feature Syndicate.)

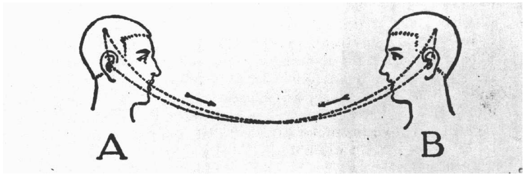
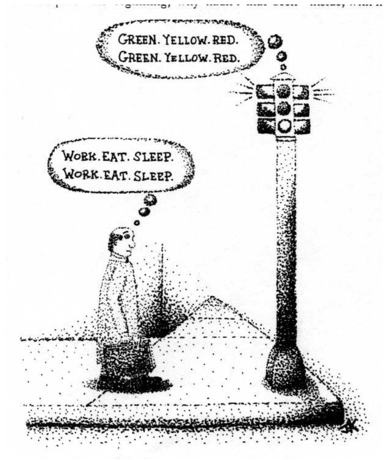


Figure 6: My ideas travel from my mind, to my mouth, to your ears, to your mind, and conversely. (From Saussure 1959: 11.)

Figure 7: How misunderstanding can arise.

Figure 8: On conveying meaning.



A *New Yorker* cartoon illustrating Winston's Problem.

"How birds see the world." (A *Far Side* cartoon illustrating Winston's Problem.)

Figure 9: