David. (1995d). Powerful evidence, but he's still O. J. *New York Times,* July 8.
David. (1995e). Jury clears Simpson in double murder; spellbound nation
es on verdict. *New York Times,* October 4, 1.
resa, Walker, Thaai & Minton, Torri. (1995). What verdict means for blacks.
Francisco Chronicle October 6, 1.
ireya. (1995). Influence of Simpson trial worries judges and lawyers. *New Times,* May 29.
ieth. (1995). One hateful word. *New York Times,* March 19.
mpson case: A legal aberration (unsigned editorial). (1995). *San Francisco nicle,* October 7.
nnis J. (1996). Supersleuths' offer to O.J. *San Francisco Examiner* May 26, 1.
arry. (1995). Judgment Day. *Newsweek,* October 9, 25-33.
. (1995). The L.A. shock treatment. *New York Times,* October 4.
. (1995). A bit reluctantly, a nation succumbs to a trial's spell. *New York s,* February 7.
A.M. (1995). On my mind: Verdict on a trial. *New York Times,* October 6.
iam. (1995). Essay: After the aftermath. *New York Times,* October 12.
n verdict (unsigned editorial). (1995). *The New York Times,* October 4.
frey. (1995). A horrible human event. *The New Yorker,* October 23, 40-49.
tt. (1995). Simpson prosecutors pay for their blunders. *New York Times,* ber 4.

Dogmas of Understanding

HERBERT H. CLARK
Stanford University

Investigators of language understanding have made a number of idealizations in order to study it, but many of these idealizations have turned into dogmas—convictions that are impervious to evidence. Because of these dogmas, investigators have often ignored, dismissed, or ruled out of court common features of everyday language such as indirect meaning, word innovation, phrasal utterances, interjections, listener roles, listener background, specialized lexicons, joint actions by speakers and addressees, disfluencies, changes of mind, gestures, eye gaze, pretense, and quotations. I describe eleven common dogmas of understanding, some evidence against them, and some of the dangers they pose for the study of understanding. Using language is fundamentally social, I argue, and social features appear to influence understanding at many, perhaps most, levels of processing.

What is language understanding? Those of us who study it have surprisingly different answers, and we cannot all be right. Many conceptions of understanding are incompatible with each other, and many, too, are incompatible with the facts. Yet it is these conceptions that guide us in developing our theories and doing our research.

Most conceptions of understanding are really *pre*conceptions borrowed from linguistics, philosophy, or experimental paradigms. If I have a favorite model of syntax, or speech acts, or the lexicon, it will suggest a conception of understanding that I will adopt along with all of its premises. Even if I realize that some of its premises don't fit the facts, I may still adopt them as *idealizations*, arguing that I need them to get on with my work. It is impossible to study language without such premises and idealizations.

Yet it is all too easy for premises and idealizations like these to turn into dogmas—into convictions that are impervious to evidence. And once they become

Direct all correspondence to: Herbert H. Clark, Department of Psychology, Building 420, Stanford University, Stanford CA 94305-2130 <herb@psych.stanford.edu>.

Discourse Processes, 23, 567–598 (1997)
ISSN 0163-853X

dogmas, they become roadblocks to scientific progress. Current dogmas about language understanding, for example, have led us to ignore, dismiss, or rule out of court a great many features of everyday language—indirect meaning, word innovation, phrasal utterances, interjections, listener roles, listener background, specialized lexicons, joint actions by speakers and addressees, disfluencies, changes of mind, gestures, eye gaze, pretense, quotations, and many other phenomena. Ignoring these phenomena have, in turn, led to misguided theories and research. It is important to reflect on the current dogmas and their consequences.

In this article I examine eleven common dogmas of understanding. Three of them have primarily to do with meaning, three with the audience, three with the interaction between speakers and addressees, and two with modes of thinking. Even if we think of these as idealizations, I argue, they have outlived their usefulness. My goal is to broaden our conception of understanding and bring it more into line with the way language works in nature. I have held each of these dogmas myself at one time or another, so my presentation is heavily autobiographical.

MEANING

When people try to understand what others say, it is generally assumed that they create mental representations along the way. These may be representations of sentence meaning, utterance meaning, speaker's meaning, word meaning, and features of the immediate context. The three dogmas I examine first all reflect preconceptions about such representations.

Sentence Comprehension

In the analysis of language, sentences need to be distinguished from utterances (Grice, 1968; Lyons, 1977). The sentence *I'm hot* is a linguistic type, consisting of the words *I*, *am*, and *hot* in a particular syntactic arrangement. Its meaning, "The person uttering these sounds is, at the moment of utterance, heated up or sweaty or spicy or sexually aroused or radioactive or ... or unusually lucky," depends only on the conventional meanings of the words and the logic dictated by their syntax. It does *not* depend on when, where, why, and by whom it was produced. In contrast, the utterance "I'm hot" *is* tied to when, where, why, and by whom it was produced. When I produced it at noon on July 4 in a card game with my son, it meant "Clark is, at noon on July 4, unusually lucky." And in producing it, I meant "At noon on July 4, Clark is warning his son that he is feeling unusually lucky." In short, we can distinguish three representations for my utterance:

(A) *Sentence Meaning*: "The person uttering these sounds is, at the moment of utterance, heated up or sweaty or spicy or sexually aroused or radioactive or ... or unusually lucky"
(B) *Utterance Meaning*: "Clark is, at noon on July 4, unusually lucky"
(C) *Speaker's Meaning*: "At noon on July 4, Clark is warning his son that he is feeling unusually lucky."

In everyday life, we try to understand what speakers mean, or what utterances mean, and that is very different from comprehending what sentences mean. One of the oldest dogmas about understanding is that you have to create a sentence meaning *in order to* create an utterance or speaker's meaning:

D1: *Dogma of Sentence Meaning: For listeners to understand what a speaker means, they must first determine the meaning of the sentence uttered.*

This dogma is tacitly held by investigators in areas that range from parsing, reading, and text understanding to truth functional semantics.

The problem with the dogma is that listeners understand many utterances even though they couldn't have comprehended the sentence uttered. One piece of evidence comes from *hidden solecisms*. Consider a sign discovered by Wason and Reich (1979) on the wall of a London hospital:

1. No head injury is too trivial to ignore.

As a sentence, 1 means, "There is no head injury that is so trivial that it cannot be ignored," and that makes no sense. The sign should have read "No head injury is too trivial to *treat*." If the hospital staff had represented the sentence meaning of 1, they should have discovered the solecism, yet for years they interpreted it as "no head injury is too trivial to treat." To do that, they must have created a plausible utterance or speaker's meaning without fully determining the sentence meaning.

London hospital workers are hardly unique. In a study by Fillenbaum (1971, 1974), university students were given utterances to paraphrase in their own words. Some of the utterances contained solecisms:

2. John dressed and had a bath.
3. John finished and wrote the article on the weekend
4. Don't print that or I'll sue you.

The students normalized these solecisms over 60% of the time, as in this paraphrase for 4: "If you print that, I'll sue you." Even when they were asked to look

carefully at each paraphrase and say if there was any "shred of difference" between the original and their paraphrase, they still said no 53% of the time (see also Gleitman and Gleitman, 1970). A related phenomenon is the so-called *Moses illusion* (Erikson & Mattson, 1981; Reder & Cleeremans, 1990). When people are asked "How many animals of each kind did Moses take on the ark?" most of them answer "Two" without noticing that it was Noah, not Moses, who had been chosen for that job.

If we understand utterances incrementally, one expression at a time, the dogma of sentence meaning must apply to each expression in turn. The assumption can be treated as a separate dogma:

D2: *Dogma of Sense Selection: Listeners determine an* enumerable *set of senses for each expression, and in understanding what a speaker means, they* select *the appropriate sense from that set.*

The idea is simple. When my son heard me say "hot," he went to his mental lexicon and found a list of conventional senses for *hot*: "heated up," "sweaty," "aroused," "radioactive," "stolen," ... and "unusually lucky." In deciding what I meant, he selected the most appropriate sense—"unusually lucky." The dogma of sense selection is held by those who hold the dogma of sentence meaning and by many others as well.

For many expressions, however, we don't merely select senses: we *create* them. Consider two extracts from a newspaper column by satirist Erma Bombeck (Clark, 1983):

5. Our *electric typewriter* got married.
6. *Stereos* are a dime a dozen.

As a sentence, 5 has an anomalous meaning: Electric typewriters don't ordinarily get married. Yet in the context of Bombeck's essay, we do more than select one of the enumerable senses of *electric typewriter*. We create the novel sense "roommate who owned an electric typewriter." Contrary to the dogma of sentence meaning, the utterance meaning "our roommate who owned an electric typewriter got married" is not based on, and cannot be derived from, the sentence meaning "the electric typewriter we own got married." In contrast, 6 has a sentence meaning that *does* make sense ("stereo systems are very common"), but that isn't the basis for what Bombeck meant. In the context of her essay, we are to create a novel sense for *stereos*, namely "roommates who own stereo systems." Examples like these undercut not only the dogma of sense selection, but the dogma of sentence meaning.

Bombeck's "electric typewriter" and "stereos" are examples of *contextual constructions*, constructions that have in principle an infinity of potential senses (Clark & Clark, 1979; Clark, 1983). English has many types of contextual constructions, some of which are illustrated in Table 1. Traditionally, for example, noun compounds like *apple-juice chair* have been assumed to have a finite set of possible senses (Lees, 1960; Li, 1971); according to Levi (1978), the number for each compound is twelve. Psychological studies of compounds have maintained the same assumption. The evidence, however, shows that people can create entirely novel senses for noun compounds (Clark, 1983; Downing, 1977; Gleitman & Gleitman, 1970; Kay & Zimmer, 1976), and they often do that as quickly as they understand conventional senses (Gerrig, 1989).

Sense creation is especially clear for contextual constructions built on proper names (Clark & Clark, 1979; Clark & Gerrig, 1983), as here:

7. You misunderstand, Hayne—when I said what we need now is a *Churchill*, I was speaking of a cigar. (William Hamilton cartoon)
8. A small boy and a girl came past close to me doing an *Indianapolis* on their tricycles. (Dick Francis's *Blood Sport*)
9. You haven't *Reno'd* until you've *Ramada'd*. (newspaper advertisement for Reno Ramada Hotel and Casino)
10. One woman to another at cocktail party: He seems very *California*, but actually he's bi-coastal. (Lorenz cartoon in the *New Yorker*)

TABLE 1
Some Types of Contextual Constructions

Contextual construction	Examples
indirect descriptions	*Stereos* are a dime a dozen. Please do a *Napoleon* for the camera. (Clark, 1978, 1983; Clark & Gerrig, 1983; Sag, 1981)
compound nouns	Sit on the *apple-juice chair*. I want a *finger cup*. (Clark, 1983; Downing, 1977; Gleitman & Gleitman, 1970; Kay & Zimmer, 1976)
denominal nouns	She's a *waller*. He's a *cupper*.
denominal verbs	She *Houdini'd* her way out of the closet. My friend *teapotted* a policeman. (Clark & Clark, 1979)
denominal adjectives	He seems very *California*.
nonpredicating adjectives	That's an *atomic* clock, not a *manual* one. (Clark, 1983)
possessives	That's *Calvin's* side of the room. Let's take *my* route. (Kay & Zimmer, 1976)

The conventional lexicon, of course, doesn't list *Churchill* and *Indianapolis* as common nouns, *Reno* and *Ramada* as verbs, or *California* as an adjective. The sentence meanings of 7 through 10 are therefore anomalous, so by the dogma of sentence meaning, we shouldn't be able to understand them. With the right background, of course, we do. What sets proper nouns apart is that they don't have genuine senses; they simply refer. So to create the novel senses for 7 through 10, we cannot build on the usual word senses—as we did for Bombeck's "electric typewriters" and "stereos." We must build on what we know about the *referents* of California, Reno, Ramada, Churchill, and Indianapolis. That is surely a different process.

Although some contextual expressions jump out at us, begging for special treatment, most are indistinguishable from expressions with conventional senses. When we read "Stereos are a dime a dozen," there is no indication that *stereo* isn't being used in a standard sense. Sense creation must be available for all utterances.

Dogma of Saying

For many investigators, the dogma of sentence meaning has been replaced by a dogma that goes like this:

D3: *Dogma of Saying: For listeners to understand what a speaker means, they must first determine what the speaker is saying.*

The dogma comes from a premise of Grice's (1975, 1978) theory of saying and implicature. He asked us to imagine A standing next to an obviously immobilized car and striking up a conversation with passerby B:

A: I am out of petrol.
B: There is a garage around the corner.

All B has *said*, in Grice's terminology, is that there is a garage, a gas station, around the corner. But he also *implicates*, again in Grice's terminology, "that the garage is, or at least may be open, etc." Grice: "In the sense in which I am using the word *say*, I intend what someone has said to be closely related to the conventional meaning of the words (the sentence) he has uttered." This includes specifying the referents, the time of utterance, the intended readings of ambiguous words, and other such things. Grice's notion of saying is close to utterance meaning, so for many investigators, the dogma might read: "For listeners to understand what a speaker means, they must first determine the utterance meaning."

In Grice's scheme, what is said is logically prior to what is implicated. But is it in practice? In Grice's own example, the word *garage* (in British English) is ambiguous between "gas station" and "parking structure." To determine what B said, A had to select "gas station." But A couldn't have selected "gas station" without determining what B was *implicating*—that B's information was relevant to A's being out of petrol. The point is easy to see by changing the first line of Grice's exchange:

A: I think I am parked in an illegal parking zone.
B: There is a garage around the corner.

This time A would take *garage* to mean "parking structure," again based on what B was implicating. Contrary to the dogma, you cannot figure out what is said (in Grice's sense) without figuring out what is implicated (Clark, 1996).

An essential feature of Grice's scheme—and the dogma of saying—is that listeners determine what is said differently from the way they "work out" what is implicated. But consider *indirect reference*. When A utters "I am out of petrol," B understands A as saying, "*My car* is out of petrol"—in which he uses *I* indirectly to refer to his car. But this is something B can only "work out" by noting that A is standing next to an immobile car and by inferring *why* A uttered those words. If A had been standing beside a lawn mower, motorcycle, or powersaw, he could have meant, "My lawn mower, or motorcycle, or powersaw is out of petrol" (see Clark, 1978, 1996; Nunberg, 1979). What is said, then, is hardly the inference-free notion it is often assumed to be. It cannot be determined in general without taking account of the speaker's intentions. This evidence also goes counter to the dogmas of sentence meaning and sense selection.

For the dogma of saying to work, we must be able to define what is said for every type of utterance. But take *phrasal utterances* (Clark, 1996). When I tell a bartender "Two dry sherries," or when I tell a taxi driver "The Mark Hopkins Hotel," I am using phrases, not sentences. Now, phrasal utterances aren't merely utterances of elliptical sentences—there is no way of establishing which sentences they could be elliptical for (see Hankamer & Sag, 1976; Wittgenstein, 1958, p. 9). If so, I am not *saying* (in Grice's sense) that I want two sherries or that I want to go to the Mark Hopkins, although this is roughly what I mean. What I mean is not based on what I said. Phrasal utterances pose a special problem for the dogma because they are so common and so easy to understand.

The problem is just as serious for *interjections* like "good-bye," "oh," or "ah!" Interjections don't combine with other words to form sentences, so they don't have literal meanings. All they have are conventional uses (Clark, 1996; Wilkins, 1992). According to the *American Heritage Dictionary*, for example, *ah* is "*used* to express various emotions, such as surprise, delight, pain, satisfaction, or dislike" (emphasis mine). So with interjections people understand what speakers

mean without determining what they said (in Grice's sense), because what is said is not defined for these utterances.

To sum up so far, pragmatic information is brought into the process of understanding very early—long before the point assumed in the dogmas of sentence meaning, sense selection, and saying. It is given priority over certain pieces of sentential logic in hidden solecisms, and it is exploited in creating senses for contextual constructions and determining what is said. However these processes work, they apply to all utterances, not just those with solecisms, contextual constructions, or anomalies in saying.

THE AUDIENCE

Many conceptions of understanding are based on simplistic assumptions about the audience. One, for example, is that all listeners work on utterances in the same way:

D4: *Dogma of Undifferentiated Hearers: Listeners go about understanding utterances in the same way regardless of their role in the discourse.*

One form of the dogma is found in philosophical accounts of language use (Bach & Harnish, 1979; Grice, 1957, 1968; Récanati, 1986; Searle, 1969, 1975). According to Searle, for example, a request is "an attempt to get H to do A," where H is "the hearer" "in the presence of whom the sentence is uttered" (p. 57). It is as if the audience always consisted solely of "the hearer," the person addressed. Likewise, in most theories of word recognition, utterance comprehension, and text processing, the recipients are called "the listener" or "the reader" as if all listeners and readers were of a single type. But this dogma is false, and it is important to see why.

Participants Roles

In conversation, speakers assign listeners to a hierarchy of *participant roles* for each communicative act (Clark & Carlson, 1982; Clark & Schaefer, 1989, 1992; Goffman, 1976):

1. Participants
   1.1. Speaker as self-monitor
   1.2. Other participants
      1.2.1. Addressees
      1.2.2. Side participants
2. Overhearers (or non-participants)
   2.1. Bystanders
   2.2. Eavesdroppers

These six roles can be illustrated for an utterance from *Hamlet*:

11. Queen Gertrude to Rosencrantz and Guildenstern in conversation with Polonius, Ophelia, and King Claudius: Did he receive you well?

In Gertrude's question, (1.1) Gertrude is a self-monitor, (1.2.1) Rosencrantz and Guildenstern are addressees ("you"), and (1.2.2) Polonius, Ophelia, and Claudius are side-participants. All the other listeners—the servants, attendants, and such like—are *not* participants in her question, but overhearers. Some are (2.1) bystanders, believing that Gertrude realizes they are listening. Others are (2.2) eavesdroppers, believing that Gertrude doesn't realize they are listening.

Speakers take different stands toward the six types of listeners. Contra Searle, Gertrude's utterance is *not* "an attempt to get [the hearer] H to do A"—to get "the hearer" to say whether Hamlet received them well. Her question is only for (1.2.1) Rosencrantz and Guildenstern. She intends (1.2.2) Polonius, Ophelia, and Claudius to understand the question, but not answer it. On the other hand, she could have any of many attitudes toward the (2.1) bystanders and (2.2) eavesdroppers. She may be indifferent toward whether they understand her, or she may try to disclose to them, conceal from them, or disguise from them what she is saying (Clark & Schaefer, 1987, 1992). Gertrude also speaks so that (1.1) she herself can monitor what she says. Speakers try to design their utterances for all the listeners they believe are or might be listening. Their utterances have a feature called *audience design* (Clark & Carlson, 1982).[1]

Many aspects of understanding are influenced by audience design (Clark & Schaefer, 1992). The participants realize that Gertrude has designed her utterance for them—that she has tried to give them *conclusive* evidence about what she means. So they expect to able to *recognize* what she means. In contrast, the overhearers realize that she is under no obligation to help them—to give them conclusive evidence. She needn't enable them to understand, for example, that "he" refers to Hamlet, or that she is being disingenuous for Polonius', Ophelia's, and Claudius' benefit. All they can do is *conjecture* about what she means—to base their inferences on *inconclusive* evidence. As for Gertrude, she already knows what she means. She listens for problems the other participants might have. Understanding takes at least three distinct forms:

(A) *Recognizing*: inferring what is meant from presumably *conclusive* evidence
(B) *Conjecturing*: drawing inferences about what is meant from presumably *inconclusive* evidence
(C) *Monitoring*: estimating how the participants and overhearers will do (A) and (B)

Different listeners proceed on different assumptions. Addressees expect to recognize, overhearers to conjecture, and speakers to monitor.

These distinctions have real consequences. As an addressee, Rosencrantz listens to Gertrude's question with an eye to replying. Should he tell the truth? Should he exaggerate? Should he decline to answer? How should he couch his response?

Gertrude: Did he receive you well?
Rosencrantz: Most like a gentleman.

As a side participant, Polonius listens simply to understand Gertrude's question. He wouldn't have been able to understand Rosencrantz' reply, "most like a gentleman," if he hadn't. At some level, preparing to produce the next action and preparing merely to understand it are distinct types of processes.

Overhearers should be at a disadvantage compared to addressees, and they are. In one study (Schober & Clark, 1989), person B, the addressee, was required to arrange a set of Tangram figures (abstract block figures of humans) at person A's instructions; A and B could talk as much as they needed to. But in the same room there was a third person O, an overhearer, whose job (unbeknownst to A and B) was to arrange the same figures simply by listening to A and B. All three began as strangers with no knowledge of the Tangram figures, and all three were separated by partitions. B was very accurate in arranging the figures, making errors only 2% of the time. Even though O heard precisely the same information, he or she made errors 15% of the time.

In a related study (Clark & Schaefer, 1987), A, B, and O were asked to arrange a set of pictures of familiar campus landmarks. This time A and B were friends whose job was to keep O from arranging the cards in the right order—to *conceal* crucial parts of what they meant from O. A and B didn't find it easy to do these two things at once. First, B made errors 8% of the time (not 2%). And second, O succeeded 47% of the time, where 12.5% was chance. Still, there was a huge difference between A's and O's success rate (92% vs. 47%). Because of audience design, therefore, addressees are in general more successful than overhearers at understanding what speakers mean. Just how much better depends on the topic and the speaker's attitude toward the overhearers.

Adequate Backgrounds

According to audience design, speakers try to engineer their utterances so that addressees and side participants can readily understand them. They base their design on the participants' having just the right background, what I will call an *adequate background* (Grice, 1957; 1968; 1975; Sperber & Wilson, 1986). Listeners without an adequate background should therefore have trouble understanding them. Many investigators, however, have assumed just the opposite:

D5: *Dogma of Inadequate Background: The process of understanding is identical in aspects* a, b, c, ... *whether the listener's background is adequate or not.*

The dogma is really a family of dogmas, one per aspect. Suppose we take *identifying phonetic segments* as one such aspect. In most studies of this process, people are required to listen to isolated utterances without any idea of what they are being used for. The assumption is that segment identification would be the same even if the subjects *did* know what they were being used for. The dogma treats listeners as if they were overhearers of a speaker indifferent to their understanding. That makes this dogma a close kin of the dogma of undifferentiated listeners.

The dogma of inadequate background is really a methodological convenience. It is difficult to study understanding in the wild, so investigators have developed a variety of laboratory techniques instead. Most of these techniques are built around contrived sentences presented to people isolated from any realistic human activity. It is easy to compose a set of words, sentences, or paragraphs with predefined properties and present them by computer, so that is how most studies get done. The participants in these studies are rarely told who is producing these words or why. They almost never have an adequate background. What difference does that make? For some processes, it makes a big difference, and for others, probably none: There is no general answer. The sobering fact is that many aspects of understanding once thought to be independent of background have been found not to be.

The point is easy to see for definite reference. When I tell my sister "George is talking to Jane" or "Duncan is home," I expect her to immediately be able to identify who George, Jane, and Duncan are, and what "home" refers to. At the same time, if you were overhearing, I wouldn't expect you to be able to do so. When speakers use definite references, they assume their addressees can immediately identify the individuals referred to—a feature their addressees count on in understanding them. Let me call this the *adequacy requirement* for definite references.

The adequacy requirement has been violated in one experiment after another in the study of understanding. Here is a small sample of utterances or sequences that have been used in laboratory experiments:

12. We checked *the picnic supplies. The beer* was warm. (Haviland & Clark, 1974)
13. *The haystack* was important because *the cloth* ripped. (Bransford & Johnson, 1973)

14. *The procedure* is quite easy. First you ... (Bransford & Johnson, 1973)
15. Rumor had it that, for years, *the government building* had been plagued with problems. *The man* was not surprised when he found several bugs in *the corner of his room*. (Swinney, 1979)
16. *The fishermen* decided to wait by *the bank*. (Simpson, 1981)
17. Who did *the janitor from the dorm* ask *the technician* to complain to? (Frazier & Clifton, 1989)
18. Except for *Tina, Lisa* was the oldest member of *the club*. (Gernsbacher, 1990)
19. *The director* and *the cameraman* were ready to start shooting when suddenly *the actress* fell from *the 14th floor*. (McKoon & Ratcliff, 1992)
20. *Walter* apologized to *Ronald this morning* because he damaged *the car* (Garnham, Traxler, Oakhill, & Gernsbacher, 1996)

All but one or two of the italicized definite references are inadequate: People in these experiments have no way of identifying their referents.

Many accounts of understanding turn on the very inadequacy of these references. In Bransford and Johnson's (1973) classic experiment, 13 was hard to remember unless it was presented with *parachute* as a prompt. But, if *the haystack* and *the cloth* had been adequate references, readers would already know that the cloth was part of a parachute, and the utterance would have been easy to remember. If indeed the cloth *was* part of a parachute. It might have been part of a tent, or shirt, or something else. It is the subject's *conjecture*—and *only* a conjecture—that the utterance is about a parachute, even when it is paired with the prompt *parachute*. The passage in 14 was also difficult to remember until subjects were told what *the procedure* referred to. Utterance 16 is ambiguous because *bank* could mean either "river side" or "financial institution." But if the reference to the bank had been adequate, the subjects would already know what *the bank* referred to, and there would have been no ambiguity. Indeed, it is only a conjecture that the bank was a river bank. It could just as well have been a savings bank—even for "the fishermen."

At least one aspect of understanding that changes when definite references are made adequate is parsing. One of the long-standing problems of parsing is how listeners resolve local ambiguities of attachment, as illustrated in (21) and (22):

21. The burglar blew open the safe with the new lock.
22. The burglar blew open the safe with the dynamite.

In (21), *with the new lock* is part of the definite reference *the safe with the new lock*, but in (22) it describes the instrument used for burgling the safe. With no other background, we are momentarily stymied by the ambiguity of *with the* in

(21) and (22), which we resolve only on reaching *new lock* or *dynamite*. If, however, we had an adequate background, we wouldn't find *with the* ambiguous at all. In an experiment by Altmann and Steedman (1988; see also Altmann, 1988; Crain & Steedman, 1985), people read either (23) or (24) followed by either (21) or (22):

23. A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock.
24. A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a strongbox with an old lock.

(23) provides an adequate background for (21) but not for (22), whereas (24) is adequate for (22) but not (21). Sure enough, people read the *with*-phrases in (21) and (22) more quickly when the background was adequate than when it was not.

Parsing is shaped by adequate background from the very beginning. In an experiment by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), people sitting at a table with objects on it were given this request:

25. Put the apple on the towel in the box.

When they had no other background, they couldn't know whether (a) there is an apple on a towel that goes into the box, or (b) there is an apple that goes onto a towel in a box. But when they could see two apples, one on a towel and another on a napkin, their background was adequate for interpretation (a), and as their eye movements showed, they made that interpretation immediately, showing no effect of the local ambiguity.

Most accounts of understanding, then, are really accounts of *overhearing*—how listeners with inadequate backgrounds conjecture about what speakers mean. Overhearers may be a legitimate object of study, but they are not addressees. Without evidence we cannot decide which processes are the same, and which are different, for two such different listeners—addressees and overhearers.

Monolithic Lexicons

Most accounts of understanding also take for granted that everyone who speaks a language such as English has the same grammar and lexicon. This, of course, is an idealization. English has hundreds of dialects, from central London ("Held in a sorter castle. Just like a horror film, wonnit?" David Lodge, *Nice Work*) to Trinidad ("Leela, is high time we realize that we living in a British

country and I think we shouldn't be shame to talk the people language good" V. S. Naipaul, *The Mystic Masseur*). Each of these has a different grammar and lexicon. The assumption is that we can ignore these differences in modeling understanding.

But can we? Let us consider a part of the idealization that deals with the lexicon. The dogma is this:

D6: *Dogma of the Monolithic Lexicon: Speakers and their addressees rely on a single monolithic lexicon.*

The idea is that all English speakers have essentially the same mental lexicons. You and I both have a lexical entry that pairs the phonological shape /dôg/ with the meaning "canine animal." For convenience, let me represent this pairing as [*dog*, "canine animal"], that is, as [word-form, meaning]. So when I say "Look at the dog," you access, say, the three lexical entries [*dog*, "canine animal"], [*dog*, "contemptible person"], and [*dog*, "inferior product"] and choose one. But you and I have different vocabularies—different mental lexicons. I know words you don't, and vice versa. These differences, it is assumed, are unsystematic and of no importance to theories of understanding.

But this assumption is false. The differences in people's mental lexicons *are* systematic, and listeners pay close attention to them (Clark, 1996, 1997). Each of us belongs simultaneously to many cultural communities. Person A, for example, might be a cardiologist, a football fan, a thirty-year-old, and a San Franciscan, whereas person B is a lawyer, a baseball aficionado, a senior citizen, and a Bostonian. Associated with each of these communities is a special lexicon, a *communal lexicon*. It is taken for granted within the community of English-speaking cardiologists, for example, that all of them have lexical entries for *sclerotic*, *aorta*, *myocardial*, and *infarction*. The same cardiologists don't assume that noncardiologists know these words, and they choose their words accordingly.

There is a special problem for word-forms that are common to two communities. When my neighbor tells me, "Cold weather has a significant effect on people's mental health," I infer that he is using *significant* in the everyday sense of "important," not in the technical sense of "statistically reliable." Why? Because I don't believe he belongs to the community in which *significant* has that sense. To handle examples like this, we must assume that each conventional form-meaning pairing is *indexed* to the community in which it is conventional. Each lexical entry consists of not two, but three parts, [community: word-form, meaning], as in these examples:

[physicians: *myocardia*, "muscular tissue of the heart"]
[North Americans: *nickel*, "a U.S. coin worth five cents"]

[American football aficionados: *nickel*, "a five-man defensive formation"]
[adult English speakers: *bug*, "insect"]
[computer aficionados: *bug*, "glitch in a computer program"]

The communities indexed are based on such features as nationality, residence, occupation, employment, hobby, religion, ethnicity, clubs, subculture, age cohort, and gender.

Too many accounts of understanding presuppose that listeners are more or less alike: Understanding doesn't change with their participant roles, their background, or the communities they belong to. But it does—sometimes radically. The problem is just as serious for reading. All writing presupposes a certain class of readers. Charles Dickens presupposes nineteenth century British English readers; the *Denver Post* presupposes American English readers up on current events in Denver; *Advanced Biochemistry II* presupposes chemistry students who have already mastered *Advanced Biochemistry I*. Without special knowledge, the rest of us are overhearers and often can only guess what these writings are about. The audience cannot be ignored.

INTERACTION

Language in the wild is thoroughly social. Ann and Basil, say, engage in talk in order to accomplish things together. And while one of them is speaking, the other is making comments and gestures that influence the very course of what the first says. When language is brought into the laboratory, it is usually stripped of its social features. Subjects are shut up in sound-proof booths and forced to listen to speech through earphones or read text from computer monitors. They are helpless to influence the speaker's actions, because there *is* no true speaker. Natural and laboratory settings are as different as day and night.

Autonomous Processes

To justify isolating listeners in the laboratory, investigators have made a number of idealizations, many of which have turned into dogmas. The basic dogma is this:

D7: *Dogma of Autonomous Processes: Speaking and listening are autonomous processes.*

Suppose the speaker is Ann and the listener Basil. According to this dogma, the processes by which Ann plans, formulates, and executes her speech are the same whether she is by herself or talking to Basil. Likewise, the processes by which Basil perceives, identifies, and understands Ann's speech are the same whether

he is by himself or talking to her. The dogma may have grown out of the study of written language, where writing and reading are far apart in time and place, but it is also applied to spoken language, where speaking and listening are ordinarily simultaneous. The dogma lies behind most studies of spoken language and is taken for granted even in most theories of pragmatics.

What is wrong with the dogma? The primary setting for language—for language use, language acquisition, and language change—is conversation, where the participants coordinate at all levels of processing (Clark, 1996). Suppose Ann says to Basil, "When is it?" At the lowest level, Basil must be attending to Ann's speech as she is speaking or all is lost, and the two of them exploit a variety of visual and auditory methods for synchronizing these processes. One level up, Basil must try to identify the utterance Ann is presenting, and when either of them runs into a problem, they must fix it efficiently and to their joint satisfaction. The next level up, Basil must try to understand what Ann means by her utterance—that he is asking her when some meeting is–and that may lead to other problems and their repair. One more level up, Basil must consider taking up Ann's question—should he answer it or not, and if so, how? Ann and Basil, as the current speaker and addressee, coordinate their actions on at least the four levels in Table 2. Using language is fundamentally a joint activity—from bottom to top.

One concrete problem for Ann and Basil is how to reach joint closure at each of these levels. To do that, they must satisfy a *grounding criterion* for each joint action: They must establish the mutual belief that B has attended to, identified, understood, or considered A's actions well enough for current purposes. Satisfying this criterion is called *grounding*.

The participants in a conversation have a battery of techniques for grounding at all levels (see Clark, 1996; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). One is the use of acknowledgments, or "back channels," words like *uh huh* and *yeah* (Schegloff, 1982). Another is the immediate uptake of a proposal in the prior turn, as here (1.4.1118):[2]

TABLE 2
Four Levels of Joint Action in Conversation

Level	Speaker A's actions	Addressee B's actions
Level 4	A proposes a joint project to B	B considers taking up A's proposal
Level 3	A means something particular for B	B understands what A means
Level 2	A presents an utterance to B	B identifies A's utterance
Level 1	A executes behavior for B	B attends to A's behavior

Ann: when is it
Basil: four thirty tomorrow

By giving an immediate answer to Ann's question, Basil claims to understand it and displays what he understands it to be. Other techniques give negative evidence of understanding, as here (9.1.1133):

Ann: can I speak to Jim Johnstone please?
Basil: senior?
Ann: yes .
Basil: yes - - -

With "senior?" Basil initiates a *side sequence* to clear up an ambiguity of interpretation at level 3. He would also initiate side sequences to clear up problems of hearing or simple attention. These are just a fraction of the grounding techniques that have been documented for spontaneous conversation.

How do Basil's processes change when he is in conversation with Ann? With so little research on this question, we are left with educated guesses, though many of these guesses are compelling. Let me compare Basil, an interacting conversational partner, with Oliver, a noninteracting overhearer.

At level 1, Basil must not only attend to Ann's speech, but also show her, with eye gaze, head nods, and acknowledgments, that he *is* attending. All Oliver has to do is attend; he has no opportunity or responsibility to get Ann to slow down or start over.

At levels 2 and 3, Basil must not only identify Ann's utterance and try to understand what she means, but monitor these processes. When he succeeds, he must indicate his success, as with head nods or acknowledgments. When he doesn't succeed, he must identify the problem ("Which Jim Johnstone? Probably senior") and request a repair at the next opportunity ("Senior?"). All Oliver *can* do is try to identify Ann's utterance and form a conjecture about what she means. Basil must also be prepared to display his construal of Ann's utterance, as with "Quite right," "No kidding," a smile, a frown, or another backgrounded signal. Oliver has no such responsibility, and it would be inappropriate for him to do that.

And at level 4, Basil must begin preparing to take up Ann's proposed joint project (e.g, "When is it?") and to do so precisely at the end of her utterance ("Four thirty tomorrow"). If he is delayed, he must tell her he is delayed, as with "um—four thirty tomorrow." Oliver has no such preparations to make. So, because of audience design and the opportunities for grounding, Basil and Oliver listen according to different criteria:

C1: *Addressee Criterion*: Addressees try to understand as well as they *need to* for current purposes.

C2: *Overhearer Criterion*: Overhearers try to understand as well as they are *able to* for current purposes.

Basil is therefore engaged in processes over and above those Oliver is engaged in. Although both of them listen, Basil is responsible for signaling Ann periodically about his progress. He formulates and performs many of his signals—eye gaze, head nods, acknowledgments, and the like—with precise timing *while* he is listening. These signals constitute a separate, or *collateral*, signaling system (Clark, 1996). Oliver monitors his understanding, but he does nothing more. Surely, formulating collateral signals takes extra effort, and that itself distinguishes addressees from overhearers.

A more profound difference is that Basil can use his collateral signals to alter the way he processes Ann's utterance, whereas Oliver cannot. Consider an actual exchange in which Ann is trying to get Basil to arrange one of twelve Tangram figures in the right order (Clark & Wilkes-Gibbs, 1986):

Ann: Okay, the next one is the rabbit.
Basil: Uhhh
Ann: That's asleep, you know, it looks like it's got ears and a head pointing down?
Basil: Okay

When Ann refers to the next Tangram figure as "the rabbit," she assumes that it is complete enough to allow Basil to identify the figure without delay. But it isn't, so he requests more information with "uhhh," and Ann provides it. If Basil hadn't been able to make that request, he might have processed "the rabbit" further and found the figure. It was precisely *because* he could ask that he didn't try to do that. The potential for such a request led him to abort deeper processing. If Oliver had been overhearing, he would have had no recourse but to process as deeply as he could. The result: Interacting addressees process ambiguities, underspecification, and lack of understanding differently from noninteracting listeners.

Conversation is much like playing a duet: It requires the tight coordination of the participants at all times and at all levels. Anyone who has ever played a duet recognizes the phenomenon. You practice your part until you can play it perfectly—alone. But when you join your partner, you make mistakes you have never made before, and your playing changes. It is all so much harder. Why? In joint actions, you must monitor your partners and make what you do contingent on what they do; and you must signal them to enable them to make their actions contingent on yours. These extra actions take their toll and alter your performance. Conversation is no different.

Recited Speech

Everyday speech is full of disfluencies. When Ann starts to speak, she is still deciding what to say and how to say it. When problems arise in the process, she has to manage them while she is speaking, and that leads to disfluencies. Here is an actual utterance, with the breaks in speaking marked by curly brackets (1.3.487):

26. the one {} one of the things I {} I'm told, {} I gather from {} from Alec, {.} this morning, {u:m} that {uh} Oscar feels very sore about, {.} is, that he sees this {uh} as a breaking of the consortium,

The speaker, William, pauses twice. He inserts the fillers "u:m" (an elongated *um*) or "uh" three times. He repeats *I* ("I I'm") and *from* ("from from"). He initiates a definite description ("the one _"), changes his mind, and makes a fresh start with an indefinite description ("one of the things"). He also changes his mind about "I'm told," making a fresh start with "I gather from Alec." Maclay and Osgood (1959), in their classic study of spontaneous speech, found an average of 3.34 pauses, 3.87 fillers, 1.68 repeats, and 1.48 fresh starts every 100 words. That amounts to one disfluency every nine words. By any measure, spontaneous speech is full of disfluencies, and listeners must take account of them.

In the laboratory, almost all spoken language is *recited speech*, which has no disfluencies. Here are just a few utterances that subjects have been asked to make judgments about:

27. The church was broken into last night. Some thieves stole most of the lead off the roof. (Marslen-Wilson & Tyler, 1980)
28. Charlie will soon bend his friend into submission. (Eimas & Nygaard, 1992)
29. The city council argued the mayor's position was incorrect. (Beach, 1991)

It isn't just that these utterances lack disfluencies. They sound wooden and unnatural—quite unlike spontaneous speech. (All of these utterances, of course, also suffer from inadequacy of reference.) They would never be mistaken for the speech of someone trying to communicate—even by voice recording. They lack something in their timing, intonation, and micro-disfluencies. The problem is, the very features lacking in these utterances may be central to the processes being studied in these experiments.

Recited speech has been justified with a number of idealizations. One of the most basic is often taken as dogma:

D8: *Dogma of Fluent Speech: The processes of understanding are fundamentally designed for flawless utterances.*

Most parsers, for example, work only on flawless utterances. They would die by the fifth word of William's utterance, because the sequence "the one one of the" doesn't fit any grammatical rule of English. They would also fail on his other disfluencies: "I I'm," "I'm told I gather," "from from," "u:m," and the two "uh"s. According to the dogma, however, these failures aren't relevant. The reasoning is this: Disfluencies are mere "performance errors"—speakers don't mean anything by them. They are a nuisance and, if anything, interfere with understanding. Listeners filter them out to get at the *real* utterances being produced, and it is the real utterances that get parsed. This reasoning, however, is faulty.

Most disfluencies can be divided into two parts: (1) the *problem* the speaker is trying to deal with; and (2) the *method* by which he or she is dealing with it. We often view the visible parts of disfluencies—the pauses, fillers, repeated words, fresh starts, repairs—as the problems, but they are really the solutions. The problems are usually hidden. William, for example, starts out "the one ..." but when he sees a problem continuing that way, he signals that he is replacing these words with "one of the things." He doesn't say what the problem is. He only signals *that* he has a problem. His stop and fresh start are his way of dealing with the problem.

Speakers have systematic methods for dealing with their problems, methods that are designed to give evidence about what the speakers mean. Here are a few examples:

1. *Replacements.* When speakers replace parts of what they have already produced, they design their replacements to make clear what is being replaced with what (Levelt, 1983; Schegloff, Jefferson, & Sacks, 1977). William, for example, replaced "I" by "I'm told," which he then replaced with "I gather from Alec." In each case, he returned to *I* to mark precisely where to start the replacement. Listeners cannot parse such utterances without identifying these replacements, nor can they understand the replacements without referring to the parsing. Parsing utterances and identifying replacements are inextricably intertwined.

2. *Editing expressions.* When speakers stop or make repairs, they often use editing expressions like "I mean," "you know," "no," "that is," and "or rather." These too are part of the collateral signaling system. Speakers use them to say why they are stopping (as with "no" for errors) or to specify the relation between the previous expression and the next (as with "I mean" and "you know")

(Erman, 1987; Levelt, 1983). Listeners cannot fully determine what speakers mean without reference to these expressions *in situ*.

3. *Fillers.* When speakers anticipate a delay in speaking, they often use *uh* or *um* to warn of the delay (Clark, 1994; Smith & Clark, 1993). They use *uh* for briefer delays, and *um* for longer delays.

4. *Thiy.* The word *the* is ordinarily pronounced "thuh" (with a schwa, as in the second syllable in *sofa*). But on occasion it is pronounced "thiy" (rhyming with *see*), as in "when you come {} when you come to look at *thiy* {.} thuh literature, {- I mean you know} thuh actual statements" (1.2.229). Speakers use "thiy" to signal that they are suspending their speech immediately to deal with a problem in production (Fox Tree & Clark, 1997).

5. *Repeats.* Speakers often repeat pronouns, articles, prepositions, conjunctions, and other function words (Maclay & Osgood, 1959; Stenström, 1987). William, for example, repeated *I* and *from*. Repeats like these occur overwhelmingly at the left-most words of major constituents (Clark, 1996). Repeated words are therefore highly informative about the type and size of constituent a speaker is currently producing.

It would be surprising if listeners didn't exploit the information provided in these methods, and there is evidence that they do. First, listeners are faster at detecting words following spontaneous repeats (e.g., *heart* in "... of the of the heart ...") than at detecting the same words once the repeat had been edited out (e.g., *heart* in "... of the heart ..."). In contrast, they are slower at detecting words in fresh starts (e.g., *there's* in "... and it could even—there's a little thing ...") than at detecting the same words without the prior false start (Fox Tree, 1995). Next, when people are asked factual questions such as "In which sport is the Stanley Cup awarded?" they are more likely to begin their answer with *uh* or *um* (e.g., "(1.4 sec) um (1.0 sec) hockey") the more uncertain they are of their answer (Smith & Clark, 1993). Listeners are able to use these fillers to make accurate judgments of how certain the respondents are of their answers (Brennan & Williams, 1995). Finally, suppose people listen to one of three recorded versions of spontaneous speech: (1) the original speech, complete with "um"s and "uh"s; (2) an edited version with the "um"s and "uh"s replaced by silences; and (3) an edited version with all the silences in 2 removed. Although listeners judge the speakers in 3 the most impressive, they find those in 1 more impressive than those in 2 (Christenfeld, 1995). Their judgments were softened by "uh" and "um."

The dogma of fluent speech is therefore problematic. Disfluencies are intended to be informative, and listeners use them in determining what speakers mean. Indeed, listeners couldn't parse utterances without identifying the disfluencies, or vice versa. An analogy might help. When we listen to people speak, we see them struggle to piece together an utterance out of its parts. As we watch them choose

this part, and discard that part, we get evidence of what those parts are. When all we see is the final product—especially recited speech—it is harder to identify those parts.

Resolute Minds

People often change their minds as they go through an activity. On the way to the post office, you might run into heavy traffic on one route and try another route, but because that route too is busy, you take a third route. Speakers are no different. William intended first to refer to "one of the things I'm told" (concealing the source of his information), but changed his mind and referred to "one of the things I gather from Alec this morning" (now revealing the source of his information). Indeed, his addressee gets an even better idea of what he is thinking by noticing how he has changed his mind. Although he doesn't say why he changed his mind, he could have, with "I mean," "that is," or some other editing expression. Speakers can change their minds about everything from word choice to the direction of conversation.

Most theories of understanding assume just the opposite. One version of the assumption is dogma for much of the field:

> **D9:** *Dogma of Determinate Meaning: Speakers have a determinate meaning in mind with each utterance, and it is up to the addressees to identify that meaning.*

When Ann says "Can I speak to Jim Johnstone please?" the assumption goes, she has a particular intention in mind—say, she is asking Basil to call Jim Johnstone to the telephone. Basil is successful once he identifies that intention. The dogma is taken for granted in most theories of speech acts, most theories of pragmatics, and elsewhere. But the dogma is false. What speakers are taken to mean is, in reality, determined by the speakers and their addressees *jointly*. The claim is especially evident in two phenomena—elective construals and accepted misconstruals.

Elective Construals. With many utterances, speakers deliberately offer their addressees a choice of construals, so when addressees make their choice, they help determine what the speaker is taken to mean. To take one example (Clark, 1979), a woman named Susan telephoned a number of local restaurants and asked the manager, "Do you accept credit cards?" The managers responded in one of three ways, as represented here (with their percentage occurrence):

30. Yes, we do. (44%)
31. We accept Visa and Mastercard. (16%)
32. Yes, we accept Visa and Mastercard. (38%)

The managers who took Susan's utterance solely as a yes/no question responded with 30. Those who took it solely as a request for names of credit cards responded with 31. Those who took it as both responded with 32. Susan's utterance had two elective construals—yes/no question and request for names—and she expected her addressees to chose one or the other or both.

What Susan was taken to mean, contrary to the dogma, was not determinate. It was established by the manager choosing among the options she presented him (question, request, or both) (Clark, 1996). After he had responded with 30, for example, she couldn't have replied, "No, I meant what credit cards to you accept," because he could rightly have objected, "Then why didn't you say so." Nor after 31 could she have replied, "No, I only wanted to know if you accepted credit cards," because this time he could have objected, "But I just told you." Susan put herself in the position of being taken to mean whichever of the options the manager chose. That, in part, is what made her utterance polite. Elective construals are common in everyday language use.

Accepted Misconstruals. Other times, speakers present an utterance with one intention in mind, but when an addressee misconstrues it, they change their minds and accept the new construal. Accepted misconstruals are hard to document because they rarely become public. The example here is one I recorded myself:

> Waitress: And what would you like to drink?
> Clark: Hot tea, please. Uh, English breakfast.
> Waitress: That was Earl Grey?
> Clark: Right.

With "That was Earl Grey?" the waitress shows she hadn't fully heard "English breakfast." I could have corrected her ("No, English breakfast"), but I didn't because I was just as happy with Earl Grey. But by not correcting her, the two of us took me as having ordered Earl Grey and not English breakfast. I initially intended to be taken as meaning one thing, but I changed my mind. Speakers may accept a misconstrual because they deem it too trivial, disrupting, or embarrassing to correct. Still, once it is grounded, it is taken to be what they mean.

When speakers and addressees can interact—as in everyday conversation—their actions are interdependent at all levels. Addressees not only listen, but monitor the state of their understanding, formulate signals about that state, and produce these signals with precision timing. They probably exploit the opportunities for asking for clarification to alter the way they listen. Conversational speech is full of disfluencies, which actually help listeners parse utterances and determine what speakers mean. Listeners also recognize that speakers can change their mind and leave part of the construal of utterances to them.

MODES OF THINKING

In speaking, we try to get others to understand us by having the right thoughts. What sort of thoughts are these, and how do we get them to think them? Students of understanding have tended to focus on a narrow class of answers, often assuming that these are the only possible answers—or the only interesting ones. A number of dogmas have emerged about what listeners are supposed to think. I will consider two of them.

Symbolic Processes

When we look at people using language, we tend to look at their choice of phonology, morphology, syntax, and lexicon, for these are what distinguish one language from another. We tend to ignore the features that are common to all language use—pointing, for example. It is all too easy, then, to treat utterances solely as the phonology, morphology, syntax, and words chosen—as sequences of symbols:

D10: *Dogma of Symbolic Interpretation: The signals to be understood in language use consist of symbols.*

By *signal*, I mean any action by which we mean things for others. And by *symbol*, I mean what C.S. Peirce meant: a sign "whose representative character consists precisely in its being a rule that will determine its interpretant. All words, sentences, books, and other conventional signs are symbols" (Buchler, 1940, p. 12). When Ann tells Basil, "Can I speak to Jim Johnstone please?" according to this dogma, her utterance consists of symbols, and it is Basil's job to interpret them. The problem with the dogma is that no utterance is entirely symbolic. The dogma excludes certain features that are essential to understanding, distorting our view of the process and skewing the theories that are supposed to account for them. It is important to see why.

Signs, according to Peirce (Buchler, 1940), come in three primary types—*symbols*, *indexes*, and *icons*. (1) Symbols are associated by rule with what they signify; *dog* signifies canine mammals via a conventional rule of English that associates the two. (2) Indexes bear a *physical or causal connection* with what they signify; a weather vane's direction signifies the direction of the wind by being causally connected with that direction. And (3) icons are associated with what they signify by *perceptual resemblance*; a portrait of Henry VIII signifies Henry VIII by bearing a perceptual resemblance to him. For Peirce, signs may also be "mixed signs," mixtures of symbols, indexes, and icons.

If we follow Peirce, people have three basic methods for signaling things (Clark, 1996). (1) *Describing-as* is the use of symbols, as when Ann uses *dog* to denote canine mammals. (2) *Indicating* is the use of indexes, as when I refer to an individual car by pointing at it ("That is mine"). People can indicate a thing by means of manual gestures, body direction, eye gaze, voice source, and many other actions. And (3) *demonstrating* is the use of icons, as when I denote the length of a fish by holding my hands just so far apart ("I caught a fish this long"). A demonstration is a selective depiction of what it signifies. People can demonstrate by means of manual gestures, body movements, voice quality, and many other devices. Most signals are composites of two or more of these methods. Uttering "that" while pointing is a composite of describing-as and indicating, and uttering "this long" while gesturing is a composite of describing-as and demonstrating.

But almost all linguistic utterances are composite signals. Most occur at particular places and times, directed by particular speakers at particular addressees, and need to be *anchored* to those places, times, speakers, and addressees. Speakers accomplish this by indicating. When Ann asks Basil, "Can I speak to Jim Johnstone please?" she directs her voice at Basil to indicate who is speaker (she is) and who is addressee (Basil). She uses the timing of her utterance to indicate when is now—the time at which she wants to speak to Johnstone. Although she uses the descriptive content of her words to denote *types* of things, she needs the indexical features (her voice, the direction of her speech, the timing) to refer to each *individual thing*.

Indicating is everywhere in language use. It is needed not only for anchoring utterances, but for interpreting deictic expressions such as *I*, *you*, *here*, *there*, *this*, *that*, *now*, *yesterday*, and *next*. It is also required for all definite references, though in a more complicated way. Demonstrating is also widespread in language use. It is found in iconic gestures, which are common in conversation and often essential to understanding (Kendon, 1980; McNeill, 1992; Schegloff, 1984). It is also found in all direct quotations, which are also common in conversations and narratives (Clark & Gerrig, 1990).

The problem for theories of understanding is that describing-as, indicating, and demonstrating work by radically different processes. The differences, briefly, are these.

(A) In describing-as, speakers try to get their addressees to *represent types of things*. How? By getting them to identify symbols—words, constructions, sentences—and, via the rules associated with the symbols, what they signify. For this process to work, both speakers and addressees have to access representations of word meaning (e.g., mental lexicons) and the combinatorial rules of morphology and syntax.

(B) In indicating, in contrast, speakers try to get their addressees to *locate actual objects or events* in space and time. They try to get their addressees to attend to those objects by means of gestures and other such devices. For

this process to work, both need to consult representations of their spatial and temporal surroundings—something that plays no role in describing-as.

(C) In demonstrating, speakers get their addressees to *represent the appearances* of actual objects or events. They do this by presenting their addressees with selective depictions (e.g., iconic gestures or direct quotations) of what they are signifying. But for this process to work, both speakers and addressees need to consult their memory for appearances—what things look and sound like.

Although much is known about describing-as, little is known about indicating and demonstrating. Even less is known about how people integrate the three methods in composite signals. The challenge is real, because all linguistic signals are composites of at least describing-as and indicating, and many also include demonstrating. Until we include all three methods, our accounts of understanding will remain distorted and incomplete.

Flat Talk

Talk is usually focused on one layer of actions at a time. When Gertrude asks Rosencrantz, "Did he receive you well?" and he answers, "Most like a gentleman," they are operating within a domain in which she is talking to him about Hamlet in Denmark around 1200. The playgoers at London's Royal Court Theatre in 1980, however, were aware of a more basic layer of actions in which they were seeing the actress Maggie Smith, playing Gertrude, performing the line to Michael Redgrave, playing Rosencrantz:[3]

Layer 2: Gertrude is speaking to Rosencrantz in Denmark in 1200
Layer 1: Smith and Redgrave are performing for an audience in London in 1980

Gertrude and Rosencrantz, the characters, are aware of the actions in layer 2, but the playgoers and actors are aware of both layers of action. (I think of layer 2, like the stage *Hamlet* is performed on, as resting on top of layer 1, the plane of the audience.) The playgoers ordinarily focus on the actions in layer 2—on what is rotten in the state of Denmark. Yet, at the same time, they are aware of the actions in the theater in layer 1 and appreciate how they determine the actions in layer 2 (Bruce, 1981; Clark, 1996; Goffman, 1974; Walton, 1990).

Most students of understanding see the world as flat. They presuppose that discourse takes place at a single layer of actions:

D11: *Dogma of the Single Layer: Understanding what a speaker is doing consists of representing a single layer of actions.*

Ordinary language use, however, is replete with layering. It arises not only in plays, but movies, operas, situation comedies, novels, short stories, jokes, parodies, song lyrics, and much more. The dogma is taken for granted in almost all work on parsing, sentence processing, and reading comprehension, and even in most work on pragmatics. To illustrate the problem, I will consider two phenomena—story telling and hyperbole.

Whenever people tell stories, they create at least one layer (layer 2) on top of the basic layer of the storyteller and audience (layer 1). In this actual example, Sam is speaking to Reynard (1.1.446):

let me tell you a story, - - -
a girl went into a chemist's shop,
and asked for contraceptive tablets, - -
and he said "well I've got . all kinds, and . all prices, what do you want"

For Reynard to understand Sam, he must realize that "let me tell you a story" is at layer 1, but "a girl went into a chemist's shop" is not. He must join Sam in the pretense that, at layer 2, an eyewitness, played by Sam, is telling a companion, played by Reynard, about an actual girl going into an actual chemist's shop. And with "well I've got all kinds and all prices," the eyewitness and the companion at layer 2 are making the joint pretense that, at layer 3, the chemist is telling the girl that he has all kinds and all prices. Without these layers, Reynard might think that an actual girl walked into an actual chemist's shop, or that "I" and "you" in line 4 referred to Sam and Reynard (just as "me" and "you" did in the line 1). And Reynard must appreciate *why* Sam is creating layer 2, and *why* he has the eyewitness create layer 3 (Clark, 1996).

Layering is also needed for understanding irony, sarcasm, teasing, hyperbole, understatement, puns, and rhetorical questions (Clark, 1996). Take the hyperbole in Kate's description of a visit to a women's college (1.3.560):

31. and . um then, a bell rang, - and - millions of feet, . ran, . along corridors, you know, and then they . it all died away

There weren't, of course, *millions* of feet running along the corridors, nor is Kate saying there were. Rather, she is inviting her audience to join her in the pretense that, at layer 2, there were *actually* millions of feet running along the corridors. Why create the pretense? So the audience will get a better idea of what the sound was actually like: It was *as if* there had been millions of feet. And when Kate's audience imagines millions of feet, they will appreciate just what she is trying to tell them.

As listeners, therefore, we must be able to represent two or more worlds of actions at once. At layer 1, we keep track of the discourse we are engaged in with our interlocutor, story teller, playwright, or film-maker. At layer 2, we must imagine the actions of another world, another domain. But it isn't enough just to imagine Gertrude asking Rosencrantz about Hamlet, the girl going into a chemist's shop, or the millions of feet on the corridor. We must appreciate *why* Maggie Smith (and Shakespeare), Sam, and Kate created these pretenses. Without this appreciation we understand only half of what is going on.

CONCLUSION

Language understanding is so complex that we have had to cut it into model-sized pieces to study it. But in cutting it up we have also made a number of idealizations, and many of these have become dogmas—premises we take as gospel. I have described eleven such dogmas and some of the dangers they pose to the study of understanding.

To outsiders the science we are pursuing must look very odd indeed. It doesn't seem to be about listening in general, but about overhearing. It doesn't seem to be about everyday conversation with all its give and take, but about utterances by people who cannot see or interact with each other. It doesn't seem to be about ordinary speech with its quirky rhythms, disfluencies, and signs of planning, but about the recitations of actors who aren't using their words for anything. It doesn't seem to be about people of flesh and blood, but about beings unanchored to any situation, without faces, eyes, or hands for gesturing, without the ability to pretend, tease, or play. It seems to be about language in a vacuum.

But language doesn't occur in a vacuum. Its use is fundamentally social. The basic setting for language is face-to-face conversation in which listeners work closely with speakers at many levels in broader joint activities. Language in other settings is equally social, but with the coordination at a distance. Social features such as these appear to influence understanding at many, perhaps most, levels of processing.

Acknowledgement: I thank Eve V. Clark, Alexander C. Huk, Timothy S. Y. Paek, and Mija M. Van Der Wege for critical comments on earlier drafts of this paper. The preparation of the paper was supported in part by NSF Grants SBR-9309612 and IRI-9314967.

NOTES

1. Audience design is a generalization of Garfinkel's (1967) notion of *recipient design*, which applies only to the attempt to design utterances for addressees.

2. This and several later examples are from Svartvik and Quirk's (1980) corpus of British English conversations. In them a dash denotes "a unit pause (of one stress unit)" and a period "a brief pause (of one light foot)." Each example is marked by the line in the corpus; "(1.4.1118)" denotes conversation 1.4 line 1118.

3. I have simplified this example a bit by leaving out the layer in which Shakespeare plays his part (see Clark, 1996).

REFERENCES

Altmann, G.T.M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3*, 73-97.

Altmann, G.T.M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Austin, J.L. (1962). *How to do things with words*. Oxford: Oxford University Press.

Bach, K. & Harnish, R.M. (1979). *Linguistic communication and speech acts*. Cambridge: MIT Press.

Beach, C.M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language, 30*, 644-663.

Bransford, J.D. & Johnson, M.K. (1973). Considerations of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing* (pp. 383-438). New York: Academic Press.

Brennan, S.E. & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*(3), 383-398.

Bruce, B. (1981). A social interaction model of reading. *Discourse Processes, 4*, 273-311.

Buchler, J. (Ed.). (1940). *Philosophical writings of Peirce*. London: Routledge and Kegan Paul.

Christenfeld, N. (1995). Does it hurt to say um? *Journal of Nonverbal Behavior, 19*(3), 171-186.

Clark, E.V. & Clark, H.H. (1979). When nouns surface as verbs. *Language, 55*, 430-477.

Clark, H.H. (1978). Inferring what is meant. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language*, (pp. 295-322). London: Wiley.

Clark, H.H. (1979). Responding to indirect speech acts. *Cognitive psychology, 11*, 430-477.

Clark, H.H. (1983). Making sense of nonconsense. In G.B. Flores d'Arcais & R. Jarvella (Eds.), *The process of language understanding*, (pp. 297-331). New York: Wiley.

Clark, H.H. (1994). Managing problems in speaking. *Speech Communication, 15*, 243-250.

Clark, H.H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H.H. (1997). Communal lexicons. In K. Malmkjær & J. Williams (Eds.), *Context in language learning and language understanding*. Cambridge: Cambridge University Press.

Clark, H.H. & Brennan, S.A. (1991). Grounding in communication. In L.B. Resnick, J.M. Levine, & S.D. Teasley (Eds.), *Perspective on socially shared cognition*, (pp. 127-149). Washington, DC: APA Books.

Clark, H.H. & Carlson, T.B. (1982). Hearers and speech acts. *Language, 58*(2), 332-373.
Clark, H.H. & Gerrig, R.J. (1983). Understanding old words with new meanings. *Journal of Verbal Learning and Verbal Behavior, 22*, 591-608.
Clark, H.H. & Gerrig, R.J. (1990). Quotations as demonstrations. *Language, 66*, 764-805.
Clark, H.H. & Schaefer, E.F. (1987). Concealing one's meaning from overhearers. *Journal of Memory and Language, 26*, 209-225.
Clark, H.H. & Schaefer, E.F. (1992). Dealing with overhearers. In H.H. Clark (Ed.), *Arenas of language use* (pp. 248-297). Chicago: University of Chicago Press.
Clark, H.H. & Schaefer, E.R. (1989). Contributing to discourse. *Cognitive Science, 13*, 259-294.
Clark, H.H. & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.
Crain, S. & Steedman, M. (1985). On not being led up the garden path: The use of context in the psychological syntax processor. In D.R. Dowty, L. Kartunnen, & A.M. Zwicky (Eds.), *Natural language parsing* (pp. 320-358). Cambridge: Cambridge University Press.
Downing, P.A. (1977). On the creation and use of English compound nouns. *Language, 53*, 810-842.
Eimas, P.D. & Nygaard, L.C. (1992). Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language, 31*(3), 375-395.
Erikson, T.A. & Mattson, M.E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20*, 540-552.
Erman, B. (1987). *Pragmatic expressions in English: A study of you know, you see, and I mean in face-to-face conversation*. Stockholm, Sweden: Almqvist & Wiksell International.
Fillenbaum, S. (1971). On coping with ordered and unordered conjunctive sentences. *Journal of Experimental Psychology, 87*, 93-98.
Fillenbaum, S. (1974). Pragmatic normalization: Further results for some conjunctive and disjunctive sentences. *Journal of Experimental Psychology, 102*, 574-578.
Fox Tree, J.E. (1995). Effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709-738.
Fox Tree, J.E. & Clark, H.H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151-167.
Frazier, L. & Clifton, C., Jr. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*(2), 73-157.
Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
Garnham, A., Traxler, M., Oakhill, J. & Gernsbacher, M.A. (1996). The locus of implicit causality effects in comprehension. *Journal of Memory and Language, 35*(4), 517-543.
Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale NJ: Erlbaum.
Gerrig, R.J. (1989). The time course of sense creation. *Memory and Cognition, 17*, 194-207.

Gleitman, L. & Gleitman, H. (1970). *Phrase and paraphrase: Some innovative uses of language*. New York: Norton.
Goffman, E. (1974). *Frame analysis*. New York: Harper and Row.
Goffman, E. (1976). Replies and responses. *Language in Society, 5*, 257-313.
Grice, H.P. (1957). Meaning. *Philosophical Review, 66*, 377-388.
Grice, H.P. (1968). Utterer's meaning, sentence-meaning, and word-meaning. *Foundations of Language, 4*, 225-242.
Grice, H.P. (1975). Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics*, Vol. 3: *Speech acts* (pp. 113-128). New York: Seminar Press.
Grice, H.P. (1978). Further notes on logic and conversation. In P. Cole (Ed.), *Syntax and semantics*, Vol. 9: *Pragmatics* (pp. 113-127). New York: Academic Press.
Hankamer, J. & Sag, I.A. (1976). Deep and surface anaphora. *Linguistic Inquiry, 7*, 391-426.
Haviland, S.E. & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512-521.
Kay, P. & Zimmer, K. (1976). *On the semantics of compounds and genitives in English*. Paper presented at the Sixth annual meeting of the California Linguistics Association, San Diego CA.
Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key (Ed.), *Relationship of verbal and nonverbal communication* (pp. 207-227). Amsterdam: Mouton de Gruyter.
Lees, R.B. (1960). The grammar of English nominalizations. *International Journal of American Linguistics, 26*, Publication 12.
Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.
Levi, J. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.
Li, C.N. (1971). *Semantics and the structure of compounds in Chinese*. Unpublished Ph.D. dissertation, University of California, Berkeley.
Lyons, J. (1977). *Semantics*, Vol. 1. Cambridge: Cambridge University Press.
Maclay, H. & Osgood, C.E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19-44.
Marslen-Wilson, W.D. & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*(1), 1-71.
McKoon, G. & Ratcliff, R. (1992). Inferences during reading. *Psychological Review, 99*, 440-466.
McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy, 3*, 143-184.
Récanati, F. (1986). On defining communicative intentions. *Mind and Language, 1*(3), 213-242.
Reder, L.M. & Cleeremans, A. (1990). The role of partial matches in comprehension: The Moses illusion revisited. In A.C. Graesser & G.H. Bower (Eds.), *The psychology of learning and motivation: Inferences and text comprehension*, (pp. 233-258). San Diego CA: Academic Press.

Sag, I.A. (1981). Formal semantics and extralinguistic context. In P. Cole (Ed.), *Radical pragmatics* (pp. 273-294). New York: Academic Press.

Schegloff, E.A. (1982). Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University Roundtable on Languages and Linguistics 1981*, (pp. 71-93). Washington, D.C.: Georgetown University Press.

Schegloff, E.A. (1984). On some gestures' relation to talk. In J.M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 262-296). Cambridge: Cambridge University Press.

Schegloff, E.A., Jefferson, G. & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361-382.

Schober, M.F. & Clark, H.H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211-232.

Searle, J.R. (1969). *Speech acts*. Cambridge: Cambridge University Press.

Searle, J.R. (1975). Indirect speech acts. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics*, Vol. 3: *Speech acts*, (pp. 59-82). New York: Seminar Press.

Simpson, G.B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior, 20*, 120-136.

Smith, V.L. & Clark, H.H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*, 25-38.

Sperber, D. & Wilson, D. (1986). *Relevance*. Cambridge, MA: Harvard University Press.

Stenström, A.-B. (1987). Carry-on signals in English conversation. In W. Meijs (Ed.), *Corpus linguistics and beyond*, (pp. 87-120). Amsterdam: Rodopi.

Svartvik, J. & Quirk, R. (Eds.). (1980). *A corpus of English conversation*. Lund, Sweden: Gleerup.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-660.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.

Walton, K.L. (1990). *Mimesis as make-believe: On the foundations of the representational arts*. Cambridge, MA: Harvard University Press.

Wason, P.C. & Reich, S.S. (1979). A verbal illusion. *Quarterly Journal of Experimental Psychology, 31*, 591-597.

Wilkins, D.P. (1992). Interjections as deictics. *Journal of Pragmatics, 17*, 119-158.

Wittgenstein, L. (1958). *Philosophical investigations*, 3rd ed. (G.E.M. Anscombe, Trans.). New York: Macmillan.