

Approaches to Studying World-Situated Language Use

Bridging the Language-as-Product and Language-as-Action Traditions

edited by John C. Trueswell and Michael K. Tanenhaus

**A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England**

3 Coordination of Action and Belief in Communication

Boaz Keysar and Dale J. Barr

Conversation is a stunning human achievement; it is as easy to do as it is difficult to study. In attempting to explain such a complex activity there is a tendency to appeal to complex mechanisms. We suggest that interlocutors resort to relatively simple action routines that converge on the complex activity of a conversation.

The Problem

Fundamental to understanding how conversation works is understanding what actions participants perform when they converse (Austin 1962). When I say “you know the fellows’ names” I might be asking if you know the names, suggesting you do, requesting that you tell me the names, or asking you to leave me alone because you already know the names. Such speech acts are at the heart of the classic American comedy sketch where Costello asks Abbott about the players on a baseball team:

Costello: You know the fellows’ names?

Abbott: Yes.

Costello: Well, then who’s playing first.

Abbott: Yes.

Costello: I mean the fellow’s name on first base.

Abbott: Who.

Costello: The guy on first base.

Abbott: Who is on first base.

“Who is on first base” is a true statement because the player named “Who” is indeed playing on first, but it is also an excellent question if you do not know who is on first. In the face of such ambiguity, identifying the action a speaker is performing is a

fundamental problem for conversationalists. Even when speakers explicitly label their speech act, it does not necessarily solve the problem:

Abbott: Who is on first base.

Costello: What are you asking me for?

Abbott: I'm not asking you—I'm telling you. WHO IS ON FIRST.

Costello: I'm asking you—who's on first?

Abbott: That's the man's name!

Saying you are asking does not make it a request for information; saying you are telling does not make it an act of informing.

As an attempted solution to this problem, Clark and his colleagues proposed that language users coordinate meaning via joint action (e.g., Clark 1996; Clark and Carlson 1982; Clark and Marshall 1981; Clark and Schaefer 1989). Language users can be perceived as engaging in joint projects to accomplish goals, and the meanings of what they say are grounded in these activities. They are like dancers who coordinate their steps on the dance floor. In this dance, Costello is leading:

Costello: Have you got a first baseman on first?

Abbott: Certainly.

Costello: Then who's playing first?

Abbott: Absolutely.

Costello: (pause) When you pay off the first baseman every month, who gets the money?

Abbott: Every dollar of it. And why not, the man's entitled to it.

Costello: Who is?

Abbott: Yes.

The interesting thing about this exchange is that although the joint activity of coordinating understanding requires the speakers to work together, their beliefs set them apart. They have different beliefs regarding what their actions are about—beliefs that are not mutual.

In this chapter we consider the role of action and belief in conversation. Over the last twenty years, the theory of joint action and language use that Clark and colleagues put forth has been very influential in psychology, pragmatics, and computational linguistics. This theory assumes that mutual belief is central to almost any aspect of the conversational act. We will call this the *mutuality assumption*. Though this assumption

was challenged early on (e.g., Johnson-Laird 1982; Sperber 1982; Sperber and Wilson 1982), it has become widely accepted among researchers on language use. We present arguments and evidence against different instantiations of this assumption, and show how language users coordinate meaning in a variety of settings without the mutuality of belief.

Belief, Mutual Belief, and Basic Assumptions

It is crucial to distinguish between shared and mutual belief. When you take money out of a cash machine and put it in your pocket, you believe you have cash in your pocket. If a thief saw you do that, then he also believes that. The belief that you have cash in your pocket is now shared in the sense that you both hold it. But only if you noticed that he saw you put the money in your pocket, and that he saw that you noticed, does it become mutual belief. So sharing a belief is not sufficient to make it mutual; you both must also believe that you both believe that it is mutual. Mutual belief, mutual knowledge, and common ground all basically refer to kinds of knowledge that have this criterial attribute of mutuality—that is, of being known to be shared.

The mutuality assumption appears in different forms in the literature. It can be described as a set of strong assumptions about the role of mutuality of belief in the coordination of language use, at both the level of the dyad and the level of the language community. If one takes seriously the idea that mutual belief pervades central aspects of comprehension and production, these assumptions follow:

1. Mutual knowledge is essential for the development of conventions in language communities and underlies their use in conversation.
2. Any joint action requires mutual knowledge; therefore the coordination of meaning in conversation requires mutual knowledge.
3. Language users *should* rely on mutual knowledge to coordinate meaning.
4. Comprehension is constrained to mutual knowledge.
5. Conversation relies on “conceptual pacts,” a form of mutual agreement among the members of a dyad.

We challenge these basic assumptions and argue that interlocutors rely on simpler mechanisms, which do not require the routine use of mutual information. Instead, we suggest that language users might opportunistically use mutual knowledge to diagnose and correct coordination problems.

Mutual Knowledge and the Evolution of Conventions: The Coordination of Meaning in Language Communities

Conversation constantly relies on conventions. Any time a word is used, a convention is invoked. Conventions are arbitrary; there is no “natural” relationship between the word *dog* and its meaning. Such conventions must therefore be known by the individuals in the community; they are typically also mutually known in the community. If we mutually know that we know English, we can infer that we mutually know the conventions of the language.

It is assumed that conventions are established and sustained by mutual knowledge (Lewis 1969). It is also assumed that in order for language users to rely on a convention in conversation, they must realize that it is mutually known with their interlocutor (Clark and Marshall 1981; Clark 1996). We suggest that both assumptions are incorrect. We first present the argument concerning the establishment of conventions and later address the role of mutual knowledge during conversation.

The creation of a linguistic convention requires coordination by community members. A new word is not a convention until knowledge of the word has disseminated throughout the community. But is it necessary that community members represent it as mutually known for this process to take place? Barr (chapter 16, this volume; 2003) shows that conventions can evolve in a linguistic community even without the benefit of mutual knowledge.

Barr demonstrates this with a community of agents simulated in a computer program. The agents pair up into dyads and play a simple communication game, in which one agent attempts to communicate four meanings to an addressee. The problem is, they start with no conventional meanings. Although each agent has four forms and four meanings, every agent starts with random mappings between the forms and meanings. This should make for difficult communication, so the agents also have a very simple update function that tells them how to change their mappings based on the successes or failures of their past interactions with other members of the community (see Barr 2003 for details).

Just as with any other linguistic community where many individual conversations take place simultaneously, pairs of agents interacted in parallel during each “round,” and then broke off to seek out new partners for the next round. Agents were most likely to interact with other agents who lived in their “neighborhood,” but could also interact with agents who lived farther away.

Barr’s simulation shows that even with these limited skills, a community of 1,000 agents established conventions that mapped forms onto meanings, either in the form

of “dialects” or a single accepted system. The convergence on the conventionalized system was highly efficient; for agents who lived in a neighborhood of nine other agents, it took an average of 280 rounds to establish a single conventional system. The striking thing is that no agent had any representation of what was mutually known in their neighborhood or in the larger community. This simulation is analogous to an entire city converging on the use of a new slang term only because people use it with their neighbors. They start by using it in different ways, but over time they end up using it the same way, all without ever thinking about what other city residents believe the term means.

Of course, one might argue that when members of a community adopt a new slang term, most believe that almost everyone knows it. It is likely to be mutually known for many reasons, one of which is the mass media, which could nearly instantaneously make it so. This is true, but the fact that the evolution of a typical convention is accompanied by mutual knowledge of that convention does not mean that this higher-level mutual knowledge played any role in its establishment. It might have been simply an epiphenomenon. Barr’s simulation shows that such knowledge of mutuality is certainly not a necessary ingredient in a community converging on a common set of conventions.

Mutual Knowledge and Conversation: The Coordination of Meaning in Dyads

While mutual knowledge may not be necessary for coordination to take place in a community, perhaps it is necessary in the dyad. We consider this possibility in three different ways. First, we ask if the coordination of understanding in conversation requires mutual knowledge by virtue of requiring a joint action; next, we consider the role of mutual knowledge in the use of linguistic precedents during conversation; and finally, we address the more general question of when people rely on mutual knowledge in comprehending language.

Joint Actions in General and the Coordination of Meaning in Particular

There is no question that the coordination of meaning in conversation requires a joint action. Abbott and Costello’s dialogue will not hold if they were not collaborating to pin down who’s on first. The question is, does all such coordination, even when a bit more successful than that of Abbott and Costello, require that each participant consider what is mutually believed by both? The mutuality assumption provides several ways of answering this question, all of which result in an unequivocal yes. One in-principle answer is an implied deduction:

All joint actions require mutual knowledge.

The coordination of meaning is a joint action.

Therefore the coordination of meaning requires mutual knowledge.

We argue against the conclusion because, we suggest, the first premise is false.

Though it seems reasonable that all joint action would depend on mutual knowledge, the premise is demonstrably false. If it were true, then any agent who is unable to represent mutual information would be unable to perform joint actions. But there are plenty of examples to suggest otherwise. When you change the clothes of a newborn it is clearly an individual action. You might as well be changing the clothes of a soft-limbed doll. But parents know that as their babies develop minimal motor control over their limbs in the first few months of life, they can, if they are in a good mood, coordinate with the parent and transform the changing of the clothes into a joint activity. Babies as young as 2 months help their parents by stiffening their arm to allow the sleeve to slide over it. Babies younger than 1 year can easily give a high five, a coordination feat analogous to adults' achievement of the prototypical joint action, the handshake. Playing peek-a-boo is a joint activity, because it requires coordination to lock eye gaze at the right moment and avert it at others. While babies as young as a couple of months can coordinate a joint action, no one would suggest that they have the ability to represent mutual belief. Only around 3 years of age do toddlers develop the ability to distinguish between what they believe and what others believe (e.g., Astington, Harris, and Olson 1988; Wellman 1990). Even if this is a conservative estimate, clearly babies who are several months old cannot represent what they believe that the other person believes about what they believe to be mutual.

One might suggest that perhaps the coordination between an adult and a baby is not truly a joint action because perhaps the adult is the one doing all the coordination. The young baby might simply learn to lift up a hand to a "give me high five" cue while the adult does all the work of coordinating with that otherwise uncooperative hand. Yet there are plenty of other examples that are unlikely to be so easily explained away. For example, young babies can jointly attend to an object with an adult, establish mutual gaze, and so on (Corkum and Moore 1995; Dunham and Moore 1995). Young babies use mutual attention to aid understanding (Bruner 1978), and they even have a notion of intention (Baldwin 1991, 1995), but they can hardly be said to be able to represent mutual knowledge of the sort required by the mutuality assumption.

There are also examples of joint activities by mature agents who cannot represent, or are unlikely to have a representation of, mutual knowledge. Just as it takes two to

perform the conversational dance, it takes two male deer to lock horns in a well-coordinated, joint struggle. Any two street cats can jointly lock eye gaze in a highly coordinated joint action that requires the cooperation of both cats, though it is unlikely that cats would pass the false-belief test, a hallmark of possessing a theory of the other's mind (Perner, Leekam, and Wimmer 1987). Closer to our species, nonhuman primates who can clearly perform joint actions do not seem to have a theory of mind that distinguishes between what they believe and what others believe (Povinelli, Bering, and Giambrone 2000; Povinelli and Giambrone 2001).

With examples from young babies and animals, we can safely conclude that mutual knowledge or belief is by no means a necessary ingredient in joint action. Even fine-tuned mental operations such as engaging in joint attention need not require mutual knowledge. Therefore, there is no reason to believe that the coordination of meaning requires mutual knowledge simply because it is a joint action.

Coordinating Meaning without Mutual Knowledge

The mutuality assumption holds that mutual belief permeates every layer of the coordination of meaning. Every conversation you have in English involves the consideration that you mutually know that you know English, that you mutually know the meanings of its words, the rules of its syntax, and so on. While this might make sense, it is not necessarily true. Conversations in English could just as well be carried out without any regard to the fact that English is mutually known to the interlocutors.

An incident that one of us was once involved in illustrates this point. BK, a native speaker of Hebrew, had his arm twisted to join colleagues from the United States in a visit to the Arab market in the Old City of Jerusalem. BK attempted to blend in with his friends as an American tourist, carrying the required tourist gear and speaking only English. After a long negotiation in a carpet store that took place completely in English, the merchant said to BK, "You should come over for tea sometime." To which BK replied, unreflectively, "Thank you." This short conversational exchange is a prime example of an adjacency pair (Schegloff and Sacks 1973)—a kind of miniconversation. By the merchant making the offer and BK accepting, they successfully performed the joint action of coordinating meaning, not a minor achievement for an Israeli and a Palestinian.

Though the exchange was a success, it occurred without the benefit of mutual belief of the language it took place in. The merchant did not really say, "You should come over for tea sometime." He said "Tavo paam lishtot te," which is the same thing in Hebrew. And BK did not really say "Thank you," he said, "Toda," which is Hebrew for "thank you." Only after BK had replied did it hit home that the merchant had tricked

him into blowing his own cover. Clearly, BK participated in a short conversational exchange in Hebrew without recognizing that Hebrew was mutually known to him and the merchant—in fact, not even fully realizing that they were speaking Hebrew. The merchant succeeded in tricking BK because he suspected that BK knew Hebrew and that BK believed that he, the merchant, believed that BK did not speak Hebrew. Perhaps he also believed that such a lack of *mutual* knowledge of Hebrew was not going to stand in the way of the perfect execution of an adjacency pair. The merchant's trick was to unveil the sharing of Hebrew, make it mutual, and thereby reveal BK's true identity, but because that mutual knowledge was itself a *product* of BK's participation in the adjacency pair, it could not have been its ingredient.

Meaning Coordination and Linguistic Precedents

Mutual knowledge of the language is a function of membership in the general language community. A different source of mutual knowledge comes from the history of interaction of specific individuals (Clark and Marshall 1981). The coordination of meaning often requires much back-and-forth effort to jointly pin down meaning and reference. This process is often called *grounding* (Clark and Brennan 1991). Once something is “grounded” it could become part of the specific common ground between the interlocutors, part of their mutual knowledge. As such, it could serve as a precedent that facilitates their future coordination. The mutual-belief assumption would hold that such grounded precedents are specific to the dyads that originally participated in the grounding. We suggest that grounded precedents are an instantiation of a much more general mechanism of language use, which operates independently of the history of specific partners.

Conceptual Pacts and the Partner-Specificity View Consider the partner-specificity view in the context of naming. Any choice of words entails a particular conceptualization of a referent (Brown 1958). We can refer to this contribution as *the chapter*, suggesting that it is part of a book; *the essay*, suggesting that it is nonfiction; *the file*, conceiving of it as a digital entity; and so on. In choosing a name, we choose a particular conceptualization. The naming choice becomes grounded between interlocutors and could become part of their common ground. Brennan and Clark (1996) argue that when interlocutors ground a particular conceptualization, this becomes a partner-specific “conceptual pact” between a speaker and an addressee—that is, an agreement about a conceptualization that should not extend beyond the original dyad that grounded it.

To support this claim, Brennan and Clark presented speakers with photographs of objects and asked them to name the objects so that a listener could identify them. For example, speakers were supposed to identify one of two shoes. When the shoe was presented in the context of the other shoe, speakers' conceptual choice was to use the subordinate-level term *loafer* as opposed to the basic-level term *shoe*. When the speakers later received the same shoe without the context of another shoe, they still tended to refer to it as *the loafer*, even though there was no need to distinguish it from any other shoe. In producing this overinformative name, speakers violated Grice's (1975) maxim of quantity, which states that speakers should provide no more and no less information than is necessary to achieve their conversational goals. This is interesting because it suggests that conceptual pacts can override the conventions of cooperative conversation.

Brennan and Clark suggest that the reason speakers chose to keep using *loafer* was the establishment of a conceptual pact. After referring to the shoe as a loafer several times in the context of another shoe, they came to conceptualize that shoe at a subordinate level, causing them to continue to use the term even when it was more specific than necessary. But Brennan and Clark's claim is more sweeping than this; they argue that the reason speakers continued to use the overly specific term was that they knew they had a history of using the term with *that specific partner*. This is because according to the partner-specificity view, "Speakers choose their wording ... for the specific addressees they are now talking to" (p. 1484).

Using a clever design, Brennan and Clark suggested that if the name choice is indeed partner specific, then speakers should stop using the term with a new partner. To test this, they had speakers name the same pictures as before in a context requiring only basic-level description. Half of the speakers did this with a completely new partner while the other half continued with the same one. They found evidence that speakers who continued with the same partner were more likely to continue to use the term *loafer*, while those who went on with a new partner tended to revert to the basic-level term *shoe* with new partners. Because speakers did not have a conceptual pact with the new listener to call the shoe a *loafer*, they had to revert to the contextually appropriate term, *shoe*. This was taken to support the partner specificity of conceptual pacts.

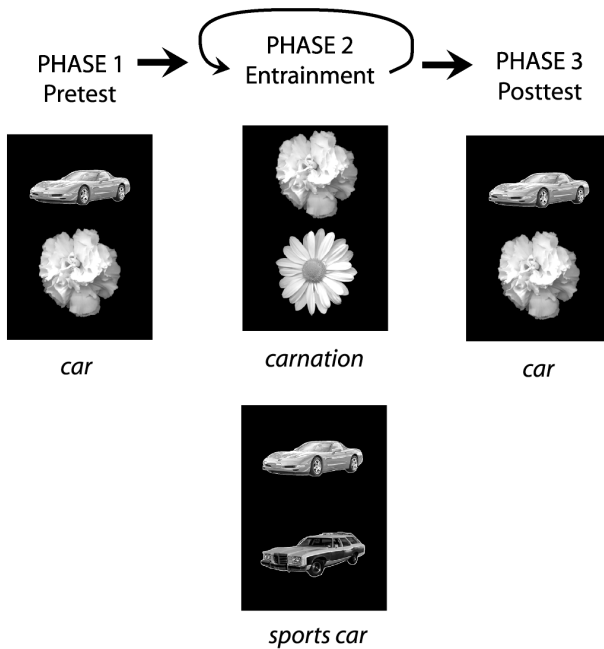
This argument reflects the mutuality assumption, that meaning coordination relies on mutual knowledge or common ground. The choice of words in reference depends on the specific history of reference between interlocutors, down to the decision to call an object either a shoe or a loafer.

However, our research suggests otherwise. We propose that word choice of the sort investigated by Brennan and Clark, while important to the coordination of meaning, does not involve partner-specific pacts or mutual expectations. Instead, the coordination of naming at this level relies on precedents that are independent of the specific individuals with whom they were established.

Grounding and Comprehension without Partner Specificity Just as the mutuality assumption predicts that speakers should choose words based on their history with the addressees they are talking to, it also predicts that addressees should develop partner-specific expectations about what speakers will say. Thus, addressees' expectations about conceptual pacts should be exclusive to the particular partner with whom the pact was established and should not generalize to new partners. We tested this prediction of the mutuality assumption in a series of experiments (Barr and Keysar 2002), one of which used the logic of Brennan and Clark (1996).

In this experiment, we presented addressees with pairs of pictures on a computer screen. Addressees heard a speaker name one of the two pictures, and selected the corresponding picture with a computer mouse. We used an eye tracker to index comprehension processes by examining the latency to fixate the target object from the onset of the referring phrase (see Eberhard et al. 1995; Tanenhaus et al. 1995).

Figure 3.1 presents the three phases of the experiment with an example item. Consider phase 1, where the addressee views a screen with a picture of a car and a flower. At this point, addressees should expect to hear the general, basic-level names *car* and *flower* because no specific precedents exist. Thus, on hearing *car* we would expect addressees to quickly fixate on the car. In phase 2, addressees saw the same two pictures, but now they were in a context where the speaker would need to use more specific names. Here, the speaker used the specific terms *carnation*, to pick out the carnation from the set of two flowers, and *sports car* to pick out the sports car from the set of two cars. These trials were repeated several times to allow addressees to entrain on these specific precedents. Now consider phase 3; this is the critical phase of the experiment. Here, the two pictures are presented back in their original context, such that speakers should, according to Grice, again use the general terms *car* and *flower*. On the other hand, listeners may still expect speakers to use the specific precedents *sports car* and *carnation*. What they actually did hear was the basic-level term *car*. If they expected to hear *carnation* instead, then they might take the sound *car* to be the beginning of the word *carnation*. Thus, expectation of the precedent should cause temporary interference as they initially consider the flower and realize that though they heard *car* . . . , . . . *nation* is not coming. The pretest, which was identical in all respects except that the

**Figure 3.1**

Design of experiment 3 in Barr and Keysar 2002.

specific precedents had not yet been established, provided a baseline to assess potential interference.

The results presented in figure 3.2 are averaged over eight items that were designed analogously to the *car/carnation* item. As the top panel of figure 3.2 clearly shows, expectation of the precedent interfered with addressees' comprehension of the basic-level term. The *x*-axis represents time from the onset of the word *car*, and the *y*-axis is the probability that the addressee is looking at a particular picture. In the pretest, for up to about 300 ms addressees were equally likely to look at the target car (represented by the black squares) or the competitor flower (the black circles), but then their eye started to favor the car over the flower. The same happened during the posttest, but the preference for the target occurred about 200 ms later. Addressees considered the carnation (white circles) longer and showed a preference for the car (white squares) later than in the pretest, suggesting that they indeed were likely to take the sound *car* as the beginning of *carnation*—that is, they expected the speaker to use the subordinate-level term.

But was this expectation partner specific, or would addressees have expected to hear *carnation* even from a speaker who had not established the precedent? To test this, half

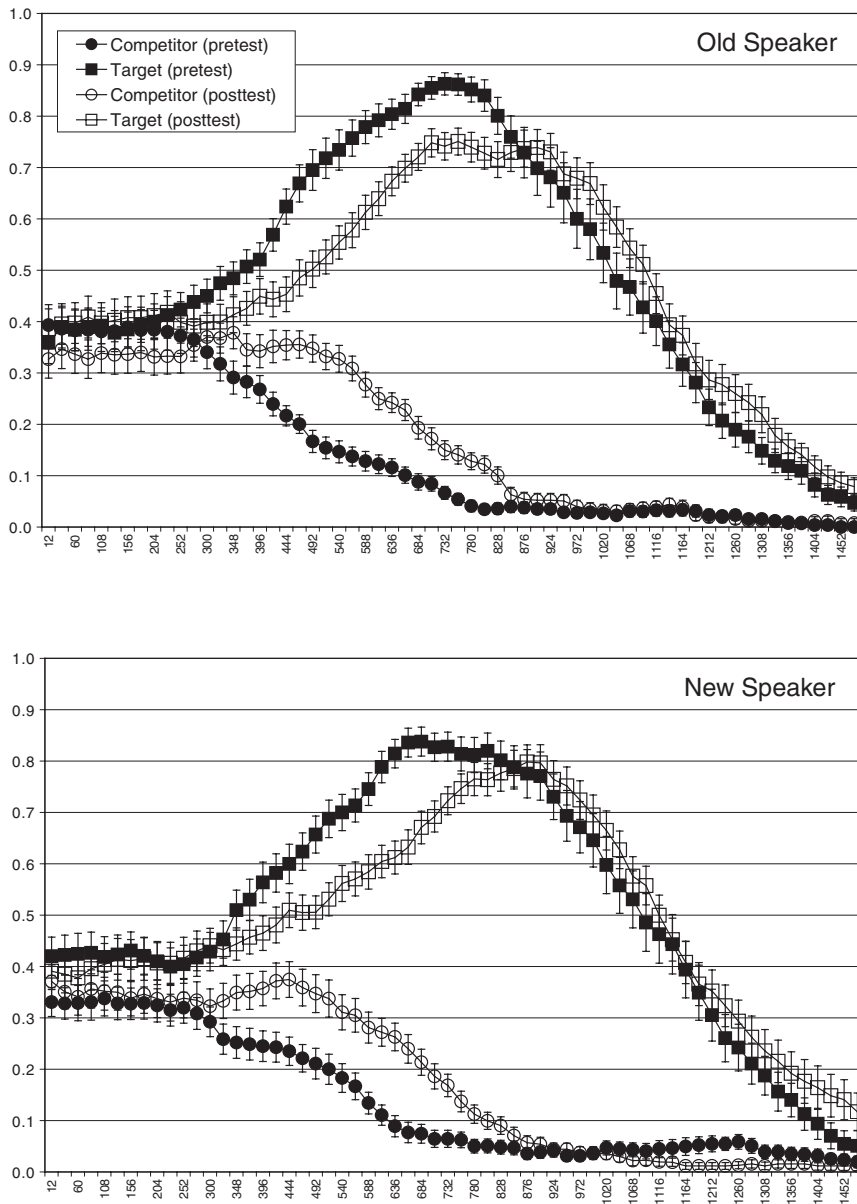


Figure 3.2

Figure 4 from Barr and Keysar 2002. The x-axis represents elapsed time from the onset of the word (0 ms). The y-axis represents the probability of fixating the different regions (see the legend).

the addressees received a completely new speaker for the posttest while the other half continued on with the same speaker. The experiment was designed so that it would seem obvious to addressees that this new speaker had never seen the pictures before and did not know the old speaker's precedents. It should be clear that according to Grice and the mutuality assumption, addressees who listen to the new speaker should have a lower expectation for the precedent *carnation* than the conventional, basic-level terms *car* or *flower*. At the very least, they should exhibit less interference from the precedent than addressees who continued on with the same person who established the precedent. But, as figure 3.2 demonstrates, on hearing *car* these addressees (bottom panel) exhibited precisely the same pattern of interference from the precedent *carnation* as addressees who continued on with the old speaker (top panel). The delay in eye movements to the car relative to the pretest was no shorter than for the old-speaker addressees. They also looked at the carnation much longer relative to the pretest. The interference caused by the precedents was not diminished, even though addressees had little reason to expect those precedents from this new speaker.

This experiment clearly shows that addressees develop strong expectations about precedents that are partner independent. These expectations are not tied to the common ground that accumulates with specific speakers.

The carnation experiment is surprising only if you truly believe in the existence of partner-specific conceptual pacts. But if you do not already believe in the theory, then your reaction is probably less enthusiastic. After all, you might think, the pictures had conventional names that are clearly known to all native English speakers. The flower is a perfectly good *flower* and also a perfectly good *carnation*; the shoe is just as much a *shoe* as it is a *loafer*. So possibly the expectations of the addressees were generalized because the choice of words reflected conventional ways of construing the referent. Perhaps if the precedents were more idiosyncratic and less conventional, addressees would have developed expectations that were partner specific.

To evaluate this possibility, consider an experiment that used nonconventional precedents (Barr and Keysar 2002, experiment 2). In that experiment we first wanted to demonstrate that grounding benefits future comprehension; that once a naming precedent had been established, the term would be more easily understood when it was used again later. We explain how we demonstrated this grounding benefit, and then describe how we evaluated whether it is partner specific.

Listeners played a referential communication game with a speaker (e.g., Glucksberg, Krauss, and Higgins 1975; Krauss and Glucksberg 1977; Wilkes-Gibbs and Clark 1992). They sat in front of a 4×4 grid of objects and the speaker (who was a confederate) gave them instructions to move objects around in the grid. We used eye movements to

the target objects as a measure of comprehension difficulty. Unlike the carnation experiment, the target objects did not have conventional names. Instead, we used objects that elicited a large variety of names from speakers in a norming study, such as a piece of paper folded in half (an “upside down V,” an “A-frame,” a “roof,” a “pyramid,” and so on). In a typical trial, a name for the object would first be grounded by the speaker, who said something like “Now put the apple above the thing that looks like a tent.” Then, after moving other objects, the speaker would say, “Now move the tent one slot down.” We measured how quickly the addressee identified the tent in the second reference, after it had been initially grounded. As a baseline we used identical trials that were not preceded by the first grounding reference.

Indeed, when the term had been grounded earlier, addressees were much faster to identify the referent than when it was not grounded—about a second and a half faster. This should not be very surprising to anyone, irrespective of their theoretical biases. If the director christened that odd object a *tent*, then next time the director referred to it, it should be easier for the listener to know what she was talking about. This illustrates the benefit of grounding for comprehension. But is this benefit specific to the grounding partner? Is it a result of a conceptual pact between the dyad that the folded paper shall be called a *tent*, especially given that someone else is very likely to conceive of the folded paper in an entirely different way (e.g., the “paper,” the “upside down V,” and so on)?

If grounding results in a partner-specific conceptual pact, then the comprehension benefit that comes with a previously grounded exchange should not fully generalize to a speaker who was not in on the original grounding. To test this, we manipulated the identity of the speaker that originally grounded the referent. In half the cases, it was the same speaker who would refer to it in the second utterance. In the other half, it was a different speaker from the one who would later refer to it in the second utterance. Surprisingly, the grounding benefit was identical regardless of whether the original grounding had been performed by the same speaker who delivered the test utterance or by another speaker who clearly did not share mutual knowledge of the precedent. Even when the person who said *tent* the second time was different from the one who said it the first time, listeners benefited equally from having heard the term previously. This suggests that listeners’ expectations of precedents that are developed in the process of grounding, such as that of calling a piece of paper a *tent*, are not partner specific but partner independent.

Malt and Sloman (2003) report converging evidence for the idea that naming precedents are generalized beyond particular partners. In their experiment, directors and matchers played a referential communication game. Directors identified pictures of

objects for matchers. For example, they referred to an object as either a tube or a container, creating a conceptual precedent. The matchers then played as directors with new matchers, and these new matchers, in turn, later played the role of directors with a different set of new matchers. Interestingly, if the original director referred to the object as a *tube*, then the final director tended to prefer *tube* over *container* as well, even though the final directors had not interacted with the original director and had not directly established any such earlier pact with their new matchers. Such carryover is consistent with our assumption that when precedents are introduced in conversation, they function in a partner-independent way. This has the added benefit of allowing a community of individuals to converge on a conventionalized form.

A strange picture emerges from experiments on grounding and conceptual pacts. On the one hand, Brennan and Clark present some evidence that speakers or directors use terms they grounded with the specific partner. On the other hand, we find that their partners do not seem to care who grounded terms with them; comprehension seems to benefit from the existence of a precedent independent of the identity of its source. This seems like an unusual asymmetry between speakers and addressees, especially given that interlocutors constantly alternate between the roles of a speaker and an addressee. Yet Brennan and Clark report that speakers' reversion to basic-level terms with new listeners was not immediate but gradual. In other words, they did not find the partner specificity they had predicted, namely, that "speakers *choose* their wording . . . for the specific addressees they are *now* talking to" (p. 1484; emphasis added), but instead they found that "the references in our task *emerged* not from solitary choices on the part of the director, but *from an interactive process* by both director and matcher" (p. 1491; emphasis added). Thus speakers initially continued using the precedents with the new addressees, who probably expressed some surprise at these overly specific terms. The speakers could have sensed the addressees' surprise, which would cause them to revert to the basic-level terms. This pattern of results for speakers seems quite consistent with our results for addressees.

The Coordination of Meaning and the Process of Comprehension

Referring in conversation is a prototypical joint action. It requires the coordination of one person to name or describe a referent and another to pick that referent out of the set of alternatives. Grounding is often part of a successful act of referring, but sometimes objects can become part of common ground in other ways. For example, when objects are mutually visible, they become part of the mutual knowledge of interlocutors (e.g., Clark 1996; Clark, Schreuder, and Buttrick 1983). When you sit at a table for dinner, the salt and pepper between you and your companion are part of your

mutual knowledge. So when you ask your companion to pass the salt, she can simply restrict her search to the objects that are mutually visible to the two of you and zoom into the intended referent. She need not consider another saltshaker that is on the counter behind you.

We conducted a series of experiments to test if meaning coordination benefits from such focus on mutually visible information. In one experiment subjects played a communication game with a partner (Keysar et al. 2000). There were objects on the table between the players, and the director, a confederate, gave the addressee instructions to move these objects around. The objects were clearly part of the common ground of the two because they were mutually visible. We distinguished between the perspectives of the two by occluding some objects so that they were invisible to the director. So while most objects were mutually visible, the identities of the occluded objects were privy only to the addressee. Those objects were obviously not part of the common ground with the director and could not have been referred to by the director. The central question was, would the addressee consider as potential referents only objects that were mutually visible? Would the process of jointly coordinating the meaning, of jointly identifying referents, be restricted to mutual knowledge, or would addressees consider the occluded referents that were not part of this set?

To test this, occluded spaces sometimes contained objects that fit speakers' descriptions of mutually visible objects. For example, the director told the matcher to "put the tape next to the truck." The corresponding grid contained both an audiotape that was mutually visible and a roll of Scotch tape that was visible only to the matcher. When searching for a referent for *tape*, would the addressee simply search for a tape exclusively among the objects that the director can see, in which case they should select the audiotape and not consider the hidden Scotch tape? To address this question, we used an eye tracker to monitor addressees' eye movements as they listened to a speaker's instructions to move objects.

Surprisingly, our results showed that addressees behaved quite egocentrically—they showed a strong tendency to consider hidden objects, such as the Scotch tape, as potential referents. In fact, a full quarter of the time they reached for it or attempted to move it. In general, we found that addressees searched among objects visible to them, even if those objects were clearly not accessible to the director, and therefore not part of common ground.

Even though addressees eventually selected the audiotape, sometimes only with the help of the director who had to correct them, the presence of hidden Scotch tape interfered with comprehension. Compared to the control condition, in which the occluded Scotch tape was replaced with an object that could not be a referent for *tape*

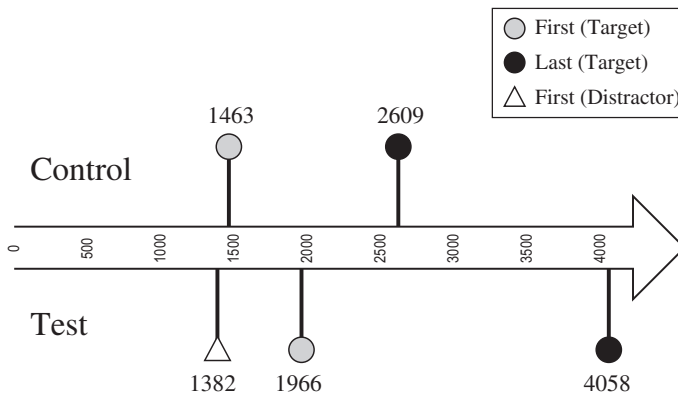


Figure 3.3

Timeline of eye fixations in experiment 1 from Keysar et al. 2000. Average latencies following the critical noun phrase (point 0). The difference between first fixation on the shared object (i.e., noticing it) and the last fixation on it (i.e., selecting it) is the decision lag. This decision took longer in the test than in the control condition.

(e.g., a toy monkey), it took subjects on average almost 2 seconds longer to select the tape as indicated by the last fixation before reaching for it (see figure 3.3). Clearly, even when people know exactly what is mutually believed, they do not restrict their search of referents to this set.

The experiment shows that a search for referents in the world is quite egocentric, in that objects visible only to the addressee are still considered referents. But this phenomenon is not unique to identifying referents in the world; it even happens when addressees search for referents in their own minds. In Keysar et al. 1998, we found that addressees who searched their memory for referents to a speaker's utterances considered information that was clearly not mutual. Just like in the tape experiment, they interpreted utterances egocentrically.

In our experiment the hidden object was always a better referent for the instructions than the target object. So while the cassette tape was a fine referent of *tape*, the hidden Scotch tape tends to be better. Consequently, even if people consider both tapes, the instruction *tape* need not be ambiguous but can yield a unique referent. This design allowed us to test a strong version of the mutuality assumption, namely, that people will spontaneously restrict comprehension to mutual information. Because the hidden referents were more typical, listeners were able to arrive at a unique referent from their own perspective, such that if they ignored mutual knowledge, they would make systematic mistakes. The design, then, allowed us to detect cases when people considered objects beyond common ground.

Although people clearly do not fully restrict comprehension to mutual knowledge, our experiment does not allow us to determine whether mutuality partially constrains comprehension from the earliest moments of comprehension. Some studies provide evidence for such a possibility (e.g., Hanna et al. 1997; Nadig and Sedivy 2002). For instance, Nadig and Sedivy (2002) found early effects of mutual knowledge in situations where an egocentric interpretation would not yield a unique referent (e.g., in response to the instruction “pick up the glass” when there were two identical glasses, one visible and one hidden). Because the ambiguity was egocentrically unresolvable, this would trigger a search for other sources of information (such as mutual knowledge) that could further circumscribe the set of potential referents. This finding is not wholly inconsistent with our view, and seems to indicate that there is some flexibility in the deployment of mutual knowledge. This raises the question of under what conditions comprehension is more or less egocentric. Our experiments have important implications for theories of language use. The prevalent mutuality assumption holds that people should restrict their comprehension to mutually known information, and that if they do not, they will make systematic errors (Clark and Carlson 1981). Subjects in the tape experiment made precisely such errors, identifying an occluded object as a referent only to later recognize their mistake and correct it. So clearly, people do not do what they “should” when comprehending language (Keysar and Barr 2002).

In fact, in contrast to the mutuality assumption, an egocentric component might not be detrimental for comprehension. Restricting comprehension to common ground would be advantageous if the most important consideration is to avoid errors. But routinely considering the other person’s perspective might be too costly for the cognitive system. Therefore, a relatively egocentric interpretation process might make more sense given that people typically have limited attentional resources (Baddeley 1986; Shiffrin 1976; Wickens 1984).

Meaning coordination could be quite compatible with a relatively egocentric comprehension process. Much of interlocutors’ own context is also part of their mutual context. So often when people use their own context they are also inadvertently using information that is mutual (Keysar 1997; Keysar, Barr, and Horton 1998). In this way, interlocutors could benefit from most of the advantages of a mutual perspective without having to consider the mutuality of the information they are using. They exploit the fact that their own perspective tends to be serendipitously mutual, and avoid the extra cognitive work involved in computing the mutuality of information.

For example, a typical conversation has a topic and an ongoing “record.” Interlocutors know both what they are talking about and what each has said so far. They know this information individually as well as mutually. But when they rely on the

conversational record, they need not consider the fact that it is mutual; they could just use it because it is known to them. The fact that a typical conversation tends to keep its topic also makes the mutuality of the topic relatively redundant. I know the topic, I do not need to also think about the fact that you know it too or that we mutually know it. It is possible that in most cases we could manage to coordinate meaning jointly with our individual knowledge.

Success and Perceived Success of Meaning Coordination

A joint action has a goal, and a central goal in conversation is the successful coordination of meaning. To increase the chances that they will be successful, interlocutors typically go back and forth in checking their understanding. Abbott and Costello realize that their joint action is incomplete as they keep coming back to the *who* in the “who’s on first.” Their exchange does not end because they realize they have not completely succeeded in coordinating their individual intentions. This illustrates the difficulty in coordinating mental entities such as beliefs or intentions.

It is easier when the goal of the enterprise is to pick out a referent in the world. When our subjects selected the hidden tape as the intended referent, they either realized their mistake because they eventually took into account the obvious occlusion, or because their partner corrected them. The success or failure of the joint action of identifying that referent can be objectively verified. This allows interlocutors to know when they have finished their joint project and can move on to the next one.

But when there is no objective evidence for the success of meaning coordination, interlocutors can only rely on their perception of its successful completion. This opens up the possibility that interlocutors can experience an illusion of mutual understanding. Here is a case where Abbott and Costello seem to experience just such an illusion:

Abbott: Well, that’s all you have to do . . .

Costello: is to throw it to first base.

Abbott: Yes.

Costello: Now who’s got it?

Abbott: Naturally.

Costello: Naturally.

Abbott: Naturally.

Costello: OK.

Abbott: Now you’ve got it.

Costello now believes that *Naturally* is the name of the first baseman, a false belief that goes undetected. Because Abbott and Costello now individually feel satisfied with their understanding, they will attempt to move on and may never realize their error.

One might want to know how frequently language users experience this illusion of mutuality. No one knows, mainly because we lack an objective way of verifying success; thus we cannot tell if what seems like successful coordination actually is. A major element in determining the prevalence of illusory coordination is how interlocutors perceive the success of their joint action. Our research suggests two tendencies that could substantially contribute to the illusion. One is the egocentric component in comprehension, which we have already discussed. The other is the tendency of speakers to overestimate their ability to convey their intentions (Keysar and Henly 2002).

Keysar and Henly gave subjects ambiguous sentences such as “Angela killed the man with the gun,” and asked them to convey one particular reading of this sentence to an addressee. Speakers were instructed to attempt to convey either that Angela used the gun to kill the man, or that she killed the man that had the gun. After the speakers said the sentence, addressees selected, from the two different readings, the meaning they thought their speakers intended. Speakers, in turn, selected the meaning they thought their addressees would select. This permitted a comparison of how well speakers gauged their addressees’ understanding.

Overall, speakers exhibited a tendency to overestimate how well their addressees understood them. While addressees were only slightly better than chance in detecting the speaker’s intention (61 percent accuracy), speakers believed that, on average, they would understand them 72 percent of the time. When addressees did not understand the speakers’ intention, in 46 percent of the cases speakers thought they had. Only 12 percent of the time did speakers think their addressees had not understood them when in fact they had. The vast majority of speakers (80 percent) systematically overestimated how well they were understood, while only 10 percent underestimated it; the remaining 10 percent were well calibrated. This study discovered speakers’ overestimation with particular kinds of syntactic and lexical ambiguities. Given the abundance of ambiguities in language, it would be interesting to see whether other ambiguities such as pragmatic ones result in a smaller tendency to overestimate or perhaps lead speakers to overestimate their effectiveness even more.

Our findings lead us to the somewhat disheartening conclusion that, if speakers overestimate how well they are understood, while addressees comprehend egocentri-

cally, many coordination problems may not be detected immediately. However, all hope may not be lost, because undetected misunderstandings have a tendency to surface over the course of the conversation. The multimodal feedback present in face-to-face situations, as well as the way conversation tends to build on itself, may provide language users with much-needed, perhaps even redundant, checks on their understanding. In fact, the interactive nature of conversation, with its abundant feedback and other checks on understanding, might in itself account for why speakers do not seem to expend valuable resources tailoring utterances to mutual knowledge, and why listeners do not always use mutual information when comprehending these utterances.

The Moral

To summarize, we have argued that while the coordination of meaning in conversation is a joint action, it does not necessarily require interlocutors to resort to their mutual beliefs. Simply because meaning coordination is a joint action does not mean that it requires mutual belief; many other types of joint actions do not. We showed that in order for a community to evolve an accepted convention, its members need not represent the convention as mutually known. We argued that the joint action of grounding and the use of the resulting conversational precedent do not necessarily involve mutual belief. Addressees seem to make use of a conversational precedent irrespective of the identity of the person who established the precedent with them. We also showed that understanding reference need not rely on mutually believed information. Instead, language users often rely on their own beliefs and their own perspective. Thus, to explain the action of meaning coordination, one always needs “belief,” but not always “mutual belief.” Interlocutors take a conversational exchange as successful not when they manage to coordinate their intentions, but when they individually believe they did. Whether they are mutually right or individually wrong is beside the point.

Note

The writing of this chapter was supported by a PHS grant R01 MH49685-06A1 to Boaz Keysar. Dale Barr was supported by a postdoctoral fellowship funded by the Beckman Foundation.

Correspondence concerning this chapter should be addressed to Boaz Keysar, Department of Psychology, The University of Chicago, 5848 South University Avenue, Chicago, IL 60637; boaz@uchicago.edu.

References

- Astington, J., Harris, P., and Olson, D., eds. 1988. *Developing Theories of Mind*. New York: Cambridge University Press.
- Austin, J. L. 1962. *How to Do Things with Words*. Oxford: Clarendon Press.
- Baddeley, A. 1986. *Working Memory*. Oxford: Clarendon Press.
- Baldwin, D. 1991. Infants' contribution to the achievement of joint reference. *Child Development*, 62, 875–890.
- Baldwin, D. 1995. Understanding the link between joint attention and language. In C. Moore and P. Dunam, eds., *Joint Attention: Its Origins and Role in Development*, 131–158. Hillsdale, NJ: Erlbaum.
- Barr, D. J. 2003. Distributed design: How communities establish shared communication systems. Unpublished manuscript.
- Barr, D. J., and Keysar, B. 2002. Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418.
- Brennan, S. E., and Clark, H. H. 1996. Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493.
- Brown, R. 1958. How shall a thing be called? *Psychological Review*, 65, 14–21.
- Bruner, J. 1978. From communication to language: A psychological perspective. In I. Markova, ed., *The Social Context of Language*, 255–287. New York: Wiley.
- Clark, H. H. 1969. Linguistic processes in deductive reasoning. *Psychological Review*, 76, 387–404.
- Clark, H. H. 1996. *Using Language*. Cambridge, UK: Cambridge University Press.
- Clark, H. H., and Brennan, S. E. 1991. Grounding in communication. In L. B. Resnick, J. Levine, and S. D. Teasley, eds., *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.
- Clark, H. H., and Carlson, T. B. 1981. Context for comprehension. In J. Long and A. Baddeley, eds., *Attention and Performance IX*, 313–330. Hillsdale, NJ: Erlbaum.
- Clark, H. H., and Carlson, T. B. 1982. Hearers and speech acts. *Language*, 58, 332–373.
- Clark, H. H., and Marshall, C. R. 1981. Definite reference and mutual knowledge. In A. H. Joshi, B. Webber, and I. A. Sag, eds., *Elements of Discourse Understanding*, 10–63. Cambridge, UK: Cambridge University Press.
- Clark, H. H., and Schaefer, E. F. 1989. Contributing to discourse. *Cognitive Science*, 13, 259–294.
- Clark, H. H., Schreuder, R., and Buttrick, S. 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, 22, 245–258.

Corkum, V., and Moore, C. 1995. Development of joint visual attention in infants. In C. Moore and P. Dunham, eds., *Joint Attention: Its Origins and Role in Development*, 61–83. Hillsdale, NJ: Erlbaum.

Dunham, P. J., and Moore, C. 1995. Current themes in research on joint attention. In C. Moore and P. Dunham, eds., *Joint Attention: Its Origins and Role in Development*, 15–28. Hillsdale, NJ: Erlbaum.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., and Tanenhaus, M. K. 1995. Eye movements as a window into real time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 121–135.

Glucksberg, S., Krauss, R. M., and Higgins, E. T. 1975. The development of referential communication skills. In F. E. Horowitz, ed., *Review of Child Development Research*, 305–345. Chicago: University of Chicago Press.

Grice, H. P. 1975. Logic and conversation. In P. Cole and J. Morgan, eds., *Syntax and Semantics 3: Speech Acts*, 41–58. New York: Academic Press.

Hanna, J. E., Trueswell, J. C., Tanenhaus, M. K., and Novick, J. M. 1997. Consulting common ground during referential interpretation. Paper presented at *the Thirty-Seventh Annual Meeting of the Psychonomic Society*, November, Philadelphia.

Johnson-Laird, P. N. 1982. Mutual ignorance: Comments on Clark and Carlson's paper. In N. V. Smith, ed., *Mutual Knowledge*. London: Academic Press.

Keysar, B. 1997. Unconfounding common ground. *Discourse Processes*, 24, 253–270.

Keysar, B., and Barr, D. J. 2002. Self anchoring in conversation: Why language users don't do what they "should." In T. Gilovich, D. W. Griffin, and D. Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*, 150–166. Cambridge, MA: Cambridge University Press.

Keysar, B., Barr, D. J., Balin, J. A., and Brauner, J. S. 2000. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Sciences*, 11, 32–38.

Keysar, B., Barr, D. J., Balin, J. A., and Paek, T. S. 1998. Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, 39, 1–20.

Keysar, B., Barr, D. J., and Horton, W. S. 1998. The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7, 46–50.

Keysar, B., and Henly, A. S. 2002. Speakers' overestimation of their effectiveness. *Psychological Science*, 13, 207–212.

Krauss, R. M., and Glucksberg, S. 1977. Social and nonsocial speech. *Scientific American*, 236, 100–105.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Malt, B. C., and Sloman, S. A. 2003. Beyond conceptual pacts: Enduring influences on lexical choice in conversation. Unpublished manuscript.

- Nadig, A. S., and Sedivy, J. C. 2002. Evidence for perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329–336.
- Perner, J., Leekam, S., and Wimmer, H. 1987. Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, 5, 125–137.
- Povinelli, D. J., Bering, J. M., and Giambrone, S. 2000. Toward a science of other minds: Escaping the argument by analogy. *Cognitive Science*, 24, 509–541.
- Povinelli, D. J., and Giambrone, S. 2001. Reasoning about beliefs: A human specialization? *Child Development*, 72, 691–695.
- Schegloff, E. A., and Sacks, H. 1973. Opening up closings. *Semiotica*, 8, 289–327.
- Shiffrin, R. M. 1976. Capacity limitations in information processing, attention, and memory. In W. K. Estes, ed., *Handbook of Learning and Cognitive Processes*, vol. 4. Hillsdale, NJ: Erlbaum.
- Sperber, D. 1982. Comments on Clark and Carlson's paper. In N. Smith, ed., *Mutual Knowledge*, 46–51. London: Academic Press.
- Sperber, D., and Wilson, D. 1982. Mutual knowledge and relevance in theories of comprehension. In N. Smith, ed., *Mutual Knowledge*, 61–87. London: Academic Press.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., and Sedivy, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Wellman, H. M. 1990. *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
- Wickens, C. D. 1984. Processing resources in attention. In R. Parasuraman and D. R. Davies, eds., *Varieties of Attention*, 63–102. Orlando, FL: Academic Press.
- Wilkes-Gibbs, D., and Clark, H. H. 1992. Coordinating beliefs in conversation. *Journal of Memory and Language*, 31, 183–194.