4

Processes of Interactive Spoken Discourse: The Role of the Partner

Michael F. Schober
New School for Social Research

Susan E. Brennan
State University of New York at Stony Brook

Conversation is different from other sorts of discourse. Perhaps because we all engage in it so often, it *seems* simple: People who speak the same language send and receive a series of messages in sequence. Actually, it isn't nearly so simple, because those messages are jointly shaped on a moment-by-moment basis (Clark, 1992, 1996, 1997; Goodwin, 1981; Krauss, 1987; Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1982). Unlike people engaged in monologue or in reading or writing text, conversationalists have the opportunity to rely on their partners in ways that structure the discourse itself.

For example, speakers rely on their partners when they solicit help in completing an utterance, as in this interchange from Wilkes-Gibbs' (1986) corpus, where B, having trouble finding a word, relies on A's completion:

A: Every time I get on the elevator when it comes from the fifth floor it has this funny smell.
B: I've been using the. . uh. .
A: stairs?
B: stairs.
A: yeah.

Speakers rely on their partners when they alter what they say midstream based on their addressee's feedback—for example, by amplifying or clarifying their utterances when their addressees appear puzzled or ask what they mean, or by cutting themselves short when their addressees already seem to understand what they are getting at. Speakers who have direct evidence of what is in their partner's visual field can refer to objects differently than speakers who don't know what their partner can see, and speakers who know their partner's cultural background or level of expertise have the opportunity to tailor their utterances more appropriately than speakers or writers who don't know their audience.

Addressees in conversation can rely on their partners, too. They can often assume that utterances are designed with their needs in mind. They can expect speakers to take all of their responses into account: clarification questions, backchannel evidence of comprehension like "uh-huh" or paralinguistic cues of uncertainty like "um," or looks of confusion or impatience. Consider this example from Wilkes-Gibbs (1986), in which B's query shapes A's continuation:

A: We really just disagree on one thing.
B: Which is . . .
A: Which is that he wants to use terms like "mind" and I do not think that they're necessary.

Rather than being passive recipients of speakers' utterances, addressees in conversation can be thought of as co-creators of the utterances (Krauss, 1987).

In this chapter, we first outline features of interactive spoken discourse that distinguish it from other forms of discourse and the research methods that have evolved to examine it. We then review some of the basic facts about production and comprehension in conversation, which show various ways that speakers and addressees take each other into account, or at least seem to. We present arguments from the discourse literature about what sorts of mental processes and representations are needed to explain these findings. Based on these arguments, we lay out a set of theoretical distinctions that need to be kept in mind in the empirical study of conversation. Finally, with these distinctions in mind, we review the burgeoning and controversial empirical literature that focuses on when and whether people adapt their discourse to each other. The strengths and weaknesses of these studies highlight the difficulty of studying conversation both naturally and with experimental control, and they make it clear that the issues are far more complex than they at first appear.

WHAT MAKES INTERACTIVE SPOKEN DISCOURSE UNIQUE?

Conversation has been argued to be the *primary* site of language use, from which all other forms of discourse, including internal speech, are in some way derivative (e.g., Bakhtin, cited in Todorov, 1984; Clark, 1996; Rommetveit, 1974; Vygotsky, 1962). Although this point is not undisputed, it seems clear that conversation differs in fundamental ways from other sorts of discourse, and thus within the field of discourse studies requires its own theoretical distinctions and methods of study.

The obvious distinguishing features are that addressees can give speakers immediate feedback and that the medium is speech rather than writing. But there are others. Under Clark and Brennan's (1991) characterization, participants in prototypical face-to-face conversations are physically *copresent* in the same environment. Face-to-face conversationalists are *visible* to each other, and so, unlike telephone users, they can make use of each other's head nods, eye gaze, and gestures as they interpret each other. The participants are *audible* to one another, and so, unlike letter writers or users of Internet chat rooms, they can make use of paralinguistic cues and the timing of each other's utterances to help understand each other. They observe each other's behaviors more or less *instantaneously*—that is, with no perceptible delay—again unlike in chat rooms or over e-mail. Unadulterated conversational speech, unlike writing or audio recordings, is *evanescent* and *without record*; utterances aren't available for replay or further inspection, except in the participants' memories. Conversationalists can produce and receive *simultaneously*, unlike two-way radio or e-mail users. Unlike writers, conversationalists form their utterances extemporaneously, and they spontaneously determine the actions they take, unlike actors performing scripted lines, survey interviewers asking scripted questions, or participants in traditional religious ceremonies. Finally, conversationalists typically speak for themselves, unlike spokespersons, simultaneous translators, actors, or authors who take a fictitious narrative stance.

A key feature of conversation is that it is far less neat than forms of discourse that allow communicators to edit or plan more carefully before they produce utterances. Consider this interchange (overlapping speech is enclosed in asterisks):

D: uh hoo the next one is um let's see . okay it's similar to the one I said had the periscope
M: yes

D: but it would be scrunched down more and the second diamond is out in front like for feet . does that make any sense . the head .
M: uh
D: it's a smaller figure
M: there isn't there is no is there let me tell *you*
D: *mm hm*
M: is the highest thing on there a diamond
D: yes
M: head
D: yes
M: and then right under the diamond head to the right is a triangle
D: uh huh
M: *okay*
D: *uh huh*
M: okay (discussed in Bortfeld et al., 2001, from Schober & Carstensen's corpus)

Conversational speech is filled with phrases of indeterminate syntax, breakoffs, repairs, disfluencies, and other evidence of speakers' planning, interpretation, and coordination problems. Such bits of evidence have been proposed to function as meaningful displays of speakers' mental states (Clark, 1994, 1996; Clark & Wasow, 1998; Fox Tree & Clark, 1997; Levelt, 1989). Regardless of whether speakers *intend* these as displays, as Clark (1994, 1996) has argued, listeners can use them to infer how confident speakers are in what they are saying (Brennan & Williams, 1995) and may even use displays such as interrupted words to infer what speakers do *not* mean (Brennan & Schober, 2001).

Conversation (face to face and on the telephone) differs from other forms of interactive discourse such as voice mail, instant messaging, or e-mail in its granularity, or potential for rapid turn interchange. Turn interchange seems to follow fairly systematic principles (Sacks et al., 1974), although the principles may vary in different cultures (e.g., Moerman, 1987; Reisman, 1974; Schiffrin, 1984; Tannen, 1984, 1994). In spoken conversation, speakers often tend to minimize overlap, perhaps because it is harder to hear when more than one speaker is talking. In other forms of interactive discourse (e.g., with shared electronic whiteboards where several remote communicators can write or draw at once), communication can take place in parallel (Whittaker, Brennan, & Clark, 1991).

Another way that spontaneous spoken discourse differs from other sorts of discourse is in the roles that participants can take (Bell, 1984; Clark 1992, 1997; Clark & Carlson, 1982; Goffman, 1976; McGregor,

1986; Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). In a two-party conversation, at any one moment one party is the *speaker* and the other is the *addressee* (except when their speech overlaps, when they momentarily play both roles simultaneously); these roles switch from utterance to utterance. A third person who is part of the conversation but isn't currently being addressed is a *side participant*. A listener who isn't part of the conversation and whose presence is known to the conversationalists is a *bystander*, and a listener whose presence is unknown to the conversationalists is an *eavesdropper*. These distinctions don't make as much sense for other forms of discourse, where participant roles need to be characterized differently. For example, even if the author of a novel could be equated with the role of speaker (although it may make more sense for the narrator to be considered the speaker, if anyone is), readers aren't quite addressees (readers can't provide immediate feedback of comprehension), nor are they exactly side participants, bystanders, or eavesdroppers (yet at some level readers are being addressed; see Gerrig, 1994).

A processing theory of conversation must go beyond modeling the participants' individual processes, integrating them into a larger whole. One explicit way to do this has been developed by Clark and Schaefer (1989), whose *contribution theory* expands on insights from investigators who pay close attention to details of interaction (e.g., Goodwin, 1981; Jefferson, 1974, 1989; Schegloff & Sacks, 1973; Schegloff, 1984, 1987, 1988), as well as insights by philosophers of language (Austin, 1962; Grice, 1975; Searle, 1969). The idea is that any utterance by a speaker in a two-party conversation is simultaneously a presentation by an individual and the start of a contribution at the level of collective action, and should be modeled accordingly. When a speaker presents new information in a turn, this doesn't guarantee that the addressee will understand what was said. The speaker can assume that what was said has been *grounded* (understood well enough for current purposes) only by getting confirmatory evidence that the addressee has accepted the utterance. Only when the speaker acknowledges that the addressee has accepted the utterance can the original presentation be assumed to have been grounded, and a contribution to have succeeded. This acknowledgment may come easily; if an addressee continues with a relevant next turn, the speaker can presume that the contribution was a success. Or it can take more effort, as when an addressee needs multiturn clarification to understand what the speaker meant.

In so doing, speakers take advantage of another feature of conversation that distinguishes it from other kinds of discourse: the ability of one partner to reduce individual processing effort by relying on the

other. Speakers in conversation have the advantage that if they have trouble thinking of a word, they can solicit their addressees to complete their utterance (Wilkes-Gibbs, 1995), as in our example at the start of this chapter. Clark and Wilkes-Gibbs (1986) proposed that speakers follow a principle of *least collaborative effort*. The idea is that speakers don't merely try to minimize their own effort in producing and comprehending, but rather try to minimize the effort they expend as a pair. This becomes especially visible in situations in which speakers and addressees' relative abilities or expertise are asymmetrical. For example, speakers who know that their partners won't easily see things from the speaker's perspective (e.g., native Long Islanders discussing train timetables with non-natives) will probably put the effort into producing utterances that are intelligible to the nonnative (see Bortfeld & Brennan, 1997; Isaacs & Clark, 1987). In the reverse situation, nonexpert speakers might rely on their expert addressees to fill in any missing information. Even though the precise nature of collaborative effort can be difficult to define in less concrete contexts (Schober, 1998a), the notion of least collaborative effort highlights the extent to which conversation is a joint activity, requiring as much coordination between both parties as dancing, shaking hands, or making music together.

RESEARCH METHODS

The two main methods for studying conversational interaction have been analyses of transcribed corpora of conversation collected outside the laboratory and laboratory studies of conversational interaction. Corpus studies have the advantage of ecological validity; in most cases, they represent conversations between unconstrained speakers who conversed for real-life purposes not set for them by an experimenter. They have the disadvantage that corpus researchers are usually overhearers; because there is no independent evidence about what speakers intended and addressees comprehended, researchers must make inferences about intentions and mental processes based purely on the transcript. Another methodological issue for corpus analyses is what is considered an appropriate sample size for generalizing about the phenomena under consideration. On one end of the spectrum, ethnographic studies by conversation analysts have contributed many important insights into the moment-by-moment processes of coordination between conversational partners. However, these results are typically based on just a handful of examples of a conversational phenomenon like third-position repair (speaker A repairs B's misunderstanding of A's original utterance) to demonstrate the existence of the

phenomenon and are less concerned with the frequency or generalizability of the phenomenon. On the other end of the spectrum, large-scale corpus analyses rely on frequency counts of the phenomena they examine (e.g., Bortfeld et al., 2001; Clark & Wasow, 1998; Fox Tree & Clark, 1997; Poesio & Vieira, 1998), which runs the risk of lumping together phenomena that aren't strictly the same.

Laboratory studies of task-oriented conversation have the advantage of allowing researchers to assess speakers' intentions and addressees' comprehension independently of the conversation, through external behaviors like grasping and moving objects. They have the disadvantage of limiting themselves to task-oriented conversation, which may differ in important ways from small talk at a cocktail party, flirting, greetings with coworkers, or shooting the breeze. And even within task-oriented paradigms, conversations are shaped differently by different tasks.

Our focus in this chapter is on laboratory studies because we believe they most directly allow investigation of the mental processes and representations underlying conversation. However, we also believe that corpus analyses are important complements to laboratory results and provide useful checks of the generalizability of laboratory results.

The most frequently used method for studying conversational process in the laboratory is some variant of the *referential communication task*, which can be traced back to Piaget's ideas in the 1920s (see Yule, 1997), but which began to be used more frequently again in the 1960s (e.g., Glucksberg et al., 1966; Krauss & Glucksberg, 1969; Krauss & Weinheimer, 1964, 1966). In the referential communication task, two people carry out a task in which one person knows information that the other needs; the task can be carried out only through conversation. The person who knows the information has variously been called sender (Krauss & Weinheimer, 1964), director (Clark & Wilkes-Gibbs, 1986), information giver (Garrod & Anderson, 1987), instructor (Lloyd, 1991), explainer (Blakar, 1984), expert (Kraut, Miller, & Siegel, 1996), and announcer (Brennan, 1995), and the other party has been called, respectively, receiver, matcher, information follower, instructee, follower, worker, or audience. The information itself varies from task to task. In some, one partner describes objects or shapes for the other, who has several objects or shapes to choose from. In others, one partner has route information on a map that the other needs—information about how to construct a model, location information on a display, or expertise on how to repair or assemble machinery. In still others, one partner may refer to actions on a videotape that the other partner has to answer questions about later.

This basic setup allows for all sorts of variations in the nature of the material to be discussed (real physical objects, photographs, objects on computer screens, maps, mazes, missing cells in a spreadsheet, actions, stories, objects with common lexicalized labels vs. difficult to describe objects), the set of alternative objects the matcher must choose from (or what Olson [1970] called the *referent array*), and whether the two parties' materials or perspectives on the materials are identical. Experimenters can vary how often the director must refer and whether the matcher and director switch roles. They can manipulate whether the matcher is another experimental participant or a confederate, whether participants are experts or novices in the task domain (Isaacs & Clark, 1987), adults or children (Glucksberg et al., 1966), strangers or acquaintances (Fussell & Krauss, 1989; Schober & Carstensen, 2002), and native or non-native speakers (Bortfeld & Brennan, 1997). They can manipulate whether speakers are in the same physical space or talk over the telephone, whether they can see each other's faces (Boyle, Anderson, & Newlands, 1994) and each others' displays (Brennan, 1990), whether they are seated close by or at a distance (Karsenty, 1999), and whether they are side to side or at some other offset (Schober, 1993, 1995).

The benefit of this sort of task over simple observation or corpus analysis is that it provides a number of potentially objective measures. Intentions are less mysterious because they are constrained by the task, and comprehension can be measured via task performance. At a coarse grain, measures include success or failure at the task and time elapsed. At a medium grain, measures include how a speaker chooses to linearize discourse entities (the assumption is that when an entity is available early in planning, it is mentioned early in the utterance; see Bock, 1995; Levelt, 1989). At a fine grain, measures include the time course of a partner's performance, such as the timing of hand movements in reaching for an object, computer mouse movements traversing a map on a screen, and eye movements. The assumption with such online measures is that what is currently reached for, moved toward, or gazed at is information about what is currently being processed or under consideration.

BASIC FINDINGS

Across a range of referential communication and production studies (for a lengthier review, see Krauss & Fussell, 1996), some phenomena are quite robust. One consistent finding is that when speakers re-refer to objects that they have discussed earlier, their references change in various ways. First, referring expressions become shorter but are just

as comprehensible by their addressees (Bortfeld & Brennan, 1997; Carroll, 1980; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Isaacs & Clark, 1987; Krauss & Weinheimer, 1964, 1966; Schober & Clark, 1989). Another regularity is that speakers tend to go from using indefinite descriptions like *a lawyer* or *some houses* to definite descriptions like *the lawyer* and *those houses* on repeated referring to the same objects (Clark & Wilkes-Gibbs, 1986; Hupet & Chantraine, 1992; Wilkes-Gibbs & Clark, 1992).

One explanation for these changes in successive referring has been that people in conversation accrue common ground or achieve joint perspectives that they rely on later. Consider the representative example in Fig. 4.1 (Brennan, 2000, from a corpus collected by Stellmann & Brennan, 1993). A and B, separated by a barrier, were trying to match sets of duplicate pictures of tangrams (abstract geometric objects). The first interchange about a given tangram took many words, several turns, and much time. Figure 4.1 (Trial 1) shows a typical first interchange; A and B produced several proposals for how to conceptualize the tangram and finally settled on one of these. After finding the right tangram, they matched (on average) 11 more and changed director/matcher roles. By the next time they referred to the tangram (in Trial

Trial 1

A: ah boy this one ah boy all right it looks kinda like, on the right top there's a square that looks diagonal

B: uh huh

A: and you have sort of another like rectangle shape, the like a triangle, angled, and on the bottom it's ah I don't know what that is, glass shaped

B: all right I think I got it

A: it's almost like a person kind of in a weird way

B: yeah like like a monk praying or something

A: right yeah good great

B: all right I got it

Trial 2

B: 9 is that monk praying

A: yup

Trial 3

A: number 3 is the monk

B: ok

FIG. 4.1. Two people in search of a perspective.

2), the process of referring had become remarkably efficient because they could rely on a conceptual precedent (see Brennan & Clark, 1996) about how to view and label it. Although this process is most evident with atypical or complex objects, it happens as well in discussions of more common objects such as postcards of landmarks (Isaacs & Clark, 1987), maps (Brennan, 1990; Wilkes-Gibbs, 1986), common objects (Bortfeld & Brennan, 1997; Brennan & Clark, 1996; Fussell & Krauss, 1992), or photographs of children (Schober & Carstensen, 2002).

A second robust finding in spoken language production studies is that repeated tokens of a word (representing *given* information) are reduced in duration compared to the word's first mention (representing *new* information; Bard et al., 2000; Fowler & Housum, 1987; McAllister et al., 1994; Samuel & Troicki, 1998). And the same word uttered under less predictable circumstances will be articulated more clearly. For example, when speakers utter the word *nine*, they articulate the word more clearly in a sentence like "the next number is *nine*" than in a sentence in which the word is highly predictable, like "a stitch in time saves *nine*" (Lieberman, 1963).

As Fig. 4.1 demonstrates, the shortening of expressions on repeated referring is striking when speakers address a partner who is allowed to interact freely. However, this phenomenon is strongly attenuated when speakers address a silent partner, an imaginary partner, or a tape recorder (Krauss & Weinheimer, 1966; Schober, 1993). This illustrates a third fundamental finding: Speaking in dialogue differs from speaking in monologue. The words produced by speakers describing tangrams in dialogue are of shorter duration than the exact same words produced in monologues, and speakers in dialogue shorten these words on second mention in dialogue far more than they shorten words in monologue (McAllister et al., 1994). And speakers describing locations to interacting partners were more likely to do so from their own perspectives (e.g., describing a location as "on the right" when this was true from their own vantage point, but not their partner's), whereas speakers with imaginary partners were more likely to take the partner's perspective (Schober, 1993).

A fourth fundamental finding is that speakers produce descriptions that appear to be tailored to particular addressees. These descriptions may be based on speakers' prior beliefs about the addressee or on judgments about the addressee's needs in the current situation (see Krauss & Fussell, 1996, for discussion). An oft-cited experiment by Kingsbury (1968) demonstrates the power of speakers' prior beliefs about addressees: Speakers on a Boston city street gave an addressee who had asked for directions longer and more detailed instructions

when the addressee appeared to be from out of town (based on dress and accent) than when the addressee appeared to be a local (see also Isaacs & Clark, 1987). In another study, speakers' gender stereotypes influenced their choices of referring expressions (Fussell & Krauss, 1992).

Choices of expressions on repeated referring appear to be based on the recent dialogue history with a particular conversational partner as well, as Brennan and Clark (1996) demonstrated. In their studies, speakers who had evolved informative referring expressions (e.g., *the man's penny loafer*) with a particular partner continued to use those expressions even when the environment changed to allow simpler expressions to be used, whereas with *new* partners, speakers referring to the same objects tended to return to the basic level terms (e.g., *shoe*). Speakers often say more than they need to, despite Grice's (1975) maxim of quantity, which suggests that cooperative speakers are only as informative as they need to be. Such choices of course may be due to the speakers' own needs; speakers may mention what is particularly salient about an object rather than only what should be needed for listeners to identify it (Ford & Olson, 1975; Mangold & Pobel, 1988) particularly if they begin describing it before observing the other objects it needs to be distinguished from (Deutsch & Pechmann, 1982; Pechmann, 1989). Yet speakers seem to rely on more than just what they themselves find salient, adding additional descriptors when referents have what Hupet, Seron, and Chantraine (1991) have called "low discriminability" (when addressees are more likely to have trouble identifying them).

The corollary to this finding is that conceptual precedents set with successive partners can lead to a surprising degree of group consensus in a short time without explicit discussion of what the conventions should be. Garrod and Doherty (1994) demonstrated that, in a group of eight pairs of speakers referring to locations in a maze, individual pairs tended to come up with fairly idiosyncratic schemes. When they subsequently switched partners to discuss similar mazes with others in the group, the schemes began to converge. Within three to five partner switches, the entire group of pairs had converged on the same scheme. This suggests that community-wide discourse conventions can arise fairly easily as local conceptual precedents are refined.

A fifth fundamental finding concerns the experience of the addressee: *Listening* is not the same as *being addressed*. In a referential communication study by Schober and Clark (1989), matchers who interacted freely with directors did better at matching tangrams than eavesdropping matchers who did the same task silently and covertly, although both kinds of matchers heard the entire conversation. Even

when another group of silent matchers did the task later by listening to a tape recording that they could stop and start at will (presumably making it easier for them to keep up with the task), they did not do as well as the matchers who could freely interact with the directors (for a convergent finding, see Kraut, Lewis, & Swezey, 1982). Addressees also comprehend speakers' utterances differently than do side participants and bystanders (Wilkes-Gibbs & Clark, 1992). And better than no interaction at all is listening to *others* interact; noninteracting matchers did better at a referential communication task when they listened to recorded dialogues than when they listened to monologues (Fox Tree, 1999).

ARGUMENTS ABOUT WHEN—AND WHETHER—ADAPTATION OCCURS

These findings show that in interactive discourse speakers often make choices that benefit listeners, as when adults and older children simplify their syntax and exaggerate their prosody for younger children (Fernald & Simon, 1984; Shatz & Gelman, 1973) or when speakers vary what they mention and how they mention it depending on what their addressees can see (Boyle et al., 1994; Brennan, 1990; Lockridge & Brennan, 2002). And listeners behave in ways that suggest that they are adapting to speakers, as when they interpret utterances differently depending on who uttered them (consider how differently one might interpret the intention underlying the question "How much do you earn per year?" if it's asked by a prospective employer, tax auditor, date, telephone market researcher, or doting relative; see Schober, 1999). But to what extent are these true adaptations, rather than simply egocentric behaviors that look like adaptations? How much adaptation is really required for successful conversational coordination, and when does the adaptation occur? Do people need elaborate mental models of their conversational partners in order to coordinate, or can they rely on less elaborate cues?

Researchers have made various arguments in the discourse literature about these questions. One argument (Clark & Marshall, 1981) arises from the proposal that the use of definite references like *the sofa* requires speakers to model what they mutually know. The idea is that for a speaker to talk about *the sofa* (as opposed to *a sofa* or *some sofa*), she must believe that her listener can determine which particular sofa she is talking about (that the listener must be currently attending to the right sofa or be able to easily figure out which sofa is the right one). Technically, this isn't sufficient; the listener must also believe that the speaker believes that the listener knows

which particular sofa is under discussion, and the speaker must believe that the listener believes that the speaker believes that the listener knows which particular sofa is being referred to, and so forth. Of course true mutual knowledge can't really be computed because it entails a recursive series of such beliefs about the partner's beliefs. In practice, Clark and Marshall's argument goes, speakers rely on assumptions that don't require infinite computation.

These assumptions about what is mutually known are based on three main kinds of information: *physical copresence*, *linguistic copresence*, and *community co-membership*. Using physical copresence, speakers and listeners can assume that the objects in their immediate perceptual field, to which they could both potentially attend (and which, in fact, they may see each other attending to), are mutually known. Based on linguistic copresence, speakers and listeners can assume that what they have said to each other in current or previous conversations can be assumed to be mutually known. And based on their assumptions about community co-membership, speakers and listeners can assume that they mutually know various facts, beliefs, and assumptions that are shared within the many communities to which they both belong. At the most generic end of the spectrum, two adults who are both native English speakers can assume that a great number of words and constructions is intelligible to each other. And at the more particular end of the spectrum, two people who have lived together for years can assume that their particular shared vacations, meals, disagreements, and private jokes are mutually known and can be referred to. In between, the many other sorts of groups that people can belong to, by choice or not—families, professions, genders, age groups, cities, nations, religions, ethnicities—can provide further bases for assuming community-based mutual knowledge.

This matters for our current purposes because it proposes that whenever a speaker does something so simple as to use the definite article *the*, she must be relying on personal or cultural common ground with a particular audience. This requires that, in Clark and Marshall's (1981) terms, "we carry around rather detailed models of people we know, especially of people we know well" (p. 55). The argument is that having common ground with a partner requires a mental representation of the partner. That is, as speakers shift from one conversational partner to another, they switch gears much as they would when turning from talking with an English speaker to talking with a French one—activating one set of information and deactivating the other set.

According to Clark and Marshall (1981), the level of detail of the information that comes into play depends on how much speakers and listeners know about each other. When strangers meet, they can make

only generic assumptions about the communities to which each other belongs, based on appearance (e.g., woman in a postal uniform, young girl holding a toy truck, man in an expensive suit with a stylish haircut) and the situations they happen to be in (e.g., behind the counter at the post office, at the playground, at a bon voyage party). When strangers meet, their introductions and greetings often help them quickly determine some basic facts about each other that help narrow down the communities to which they belong: where they are from, their professions, their interests. The longer people know each other, the more information they have to draw on about which communities they do share and don't share, and this allows them to produce and interpret utterances appropriately.

A quite different possibility for explaining the apparent adaptation between conversational partners is represented by Garrod and Anderson's (1987) proposal that speakers and listeners in conversation rely on an *output–input coordination principle*. The idea is that conversationalists don't need to have detailed models of each other in order to perform at least some kinds of speech adaptation. A speaker who uses the same words her partner did, for example, needs only to adopt the most recent successful referring expression her partner produced; output is based on the most recent input, and the dialogue history before that should not matter. Output–input coordination is an extremely local and literal form of coordination.[1] The speaker takes the lead role with the addressee as follower (see also Sperber & Wilson, 1986).

Another logical possibility is represented by Brennan and Clark's (1996) notion of conceptual pacts, or temporary, flexible agreements by partners to conceptualize an object in the same way. On this view, a speaker's initial referring expression represents a proposal for a perspective on an object, which the addressee then ratifies or revises. The initial referring expression may be marked as a proposal via hedges (e.g., "a car, *sort of* silvery purple colored"), which drop out on re-referring. Two partners mark having reached a conceptual pact by reusing the same or similar expressions. This view differs from output–input coordination in that a conceptual pact is not local; the more well established a pact is between two people (the more often they have used it to refer to an object, in Brennan and Clark's experiments), the more likely it is to persist. Unlike Clark and Marshall's view, a conceptual pact need not be based on a model of a partner's knowledge, but may be shaped and maintained by the partner's feedback.

[1]In fact, this sort of simple, local adaptation underlies some successful interactive conversational software programs (e.g., Brennan, 1988; Graesser, Person, Harter, & the Tutoring Research Group, 2001).

To what extent are speakers' choices really adjustments to others as opposed to being automatic and self-centered? Brown and Dell (1987) contributed an important insight: What looks like a partner adjustment in production doesn't necessarily reflect partner-adjusted processing. An utterance may have been produced in a way that benefited a particular addressee, but this doesn't by any means guarantee that the speaker was taking this particular addressee into account. The speaker may have planned the utterance in a way that suited the *speaker* best—that was easiest to plan and produce. The fact that the resulting utterance was easier for the addressee to comprehend doesn't demonstrate that the speaker's processing was adapted to this *particular* listener's needs. Nor does it demonstrate even that the speaker's processing was adapted to the needs of *any* listener; the form of the utterance might reflect purely egocentric processing.

The problem, as discussed by Keysar (1997), is that whatever is mutually known is also individually known; this means that when speakers or addressees seem to be acting on the basis of mutual knowledge, it is possible that they are merely acting on the basis of their own personal knowledge. These two kinds of knowledge are difficult to tease apart unless the speaker's and addressee's perspectives differ in some measurable way.

Brown and Dell (1987) also made a useful distinction between *particular-listener* and *generic-listener* adaptations. Particular-listener adaptations include any adaptation to the unique informational needs of a particular listener and so, by definition, require a model of the partner. Generic-listener adaptations are the kinds of adjustments speakers do that would benefit any listener (presumably within a given language community).[2] For example, the fact that new or unpredictable words are more clearly articulated and given or predictable words are relatively reduced benefits listeners. But speakers seem to perform this reduction even without actual addressees; this reduction could occur if the speech production system is organized in a modular fashion. For generic listener adaptations, Brown and Dell contrasted the possibility that speakers' planning processes are encapsulated from knowledge about addressees (a "modularity" hypothesis) with the possibility that speakers actually take addressees into account while planning (a "listener-knowledge" hypothesis).[3]

[2]Note that Brown and Dell's logic assumes that speakers and addressees are monolingual and speak the same language. For the many who are multilingual, the choice of what language to use in an utterance is already a partner adaptation.

[3]Grosjean (1998) proposed that bilingual speakers in conversation enter either a bilingual or monolingual *mode* that enables them to readily produce and understand code switches in conversation or else suppress the language that is not needed. Presumably this sort of adaptation would have to be driven by listener knowledge.

Another set of arguments in the discourse literature focuses less on *whether* speakers and listeners adapt to each other, but *when* they do. Keysar and colleagues proposed a *perspective adjustment* theory (Keysar, Barr, & Horton, 1998), also called *monitoring and adjustment* (Horton & Keysar, 1996) from the perspective of the speaker or *unrestricted search* (Keysar, Barr, Balin, & Paek, 1998) from the perspective of the addressee. This theory proposes that, although speakers and addressees often ultimately make the appropriate adaptations to each other, their initial processing is always egocentric, and "common ground . . . only plays a role in monitoring" (Horton & Keysar, 1996, p. 91). On this proposal, speakers monitor whether the utterances they have already designed are appropriate from their addressees' perspectives and adjust to their partners only as a repair when they notice infelicity. Similarly, addressees first interpret utterances from their own perspective and only later adjust their interpretations to match what speakers might have intended from their (speakers') perspectives.

Horton and Keysar (1996) pitted their monitoring and adjustment theory against an extreme alternative, *initial design*, in which interpretation is restricted to *only* the information that is in common ground. Their initial design alternative takes seriously suggestions like Clark and Carlson's (1981) that "when a listener tries to understand what a speaker means, the process he goes through can limit memory access to information that is common ground between the speaker and his addressees" (p. 328), and Clark, Schreuder and Buttrick's (1983) proposal that "all the information the listener should ever appeal to is the speaker's and addressee's common ground" (p. 258). Horton and Keysar's initial design formulation strips the original proposals of some of their nuances; for example, it leaves out that people recognize that speakers can misjudge common ground (Clark, Schreuder, & Buttrick, 1983) and that speakers may take shortcuts in planning in order to produce an utterance sooner (Clark & Wilkes-Gibbs, 1986). But it creates a clear and testable processing alternative (although see Polichak & Gerrig, 1998, for arguments on its implausibility).

In considering when partner models come into play, we can distinguish among these two extreme possibilities and a *dual process* account (Bard & Aylett, 2000) or what Hanna, Tanenhaus, Trueswell, and Novick (2002) called a probabilistic or *constraint-based* account. In both the Bard and Aylett and Hanna et al. approaches, adapting to a partner is not a two-stage process, first egocentric and then adjusted. Rather, alternative interpretations or distinct processes proceed in parallel, as Tanenhaus and colleagues (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) have proposed for the language processing system more generally. Bard et al. (2000) emphasized a difference between fast automatic processes and slower controlled processes in language production. Fast automatic processes, like priming, would likely be impervious to any information about addressees and "default" to entirely egocentric processing. Controlled processes, such as updating and assessing models of a conversational partner's informational needs, would proceed simultaneously but more slowly. Under time pressure (e.g., as Horton & Keysar [1996] concluded from their study), speakers and listeners may rely on just the fast, automatic processes and thereby fail to take their partners into account. As Polichak and Gerrig (1998) pointed out, common ground (including the potential for physical copresence) was proposed by Clark and Marshall (1981) to be a source of inferences about what partners know, and so it should not be surprising if such inferences require time and effort. It would be extremely unlikely (and computationally infeasible) if interlocutors routinely pre-computed all relevant information according to whether it was in common ground (Polichak & Gerrig, 1998).

The distinction between more controlled and more automatic processing raises the likelihood that different stages of language processing should be more or less amenable to a partner's influence (see Bard & Aylett, 2000). For example, if articulation is more automatic than the word selection that precedes it (see Levelt, 1989), it might be less likely to be affected by a partner's informational needs.

THEORETICAL DISTINCTIONS

These considerations show that the question of if, when, and how conversationalists take each other into account is far more complex than it at first appears. There are a number of critical distinctions to keep in mind. First, there are several different possibilities for adjustment:

1. *Specific-partner adjustments*. In speaking, such adjustments would involve making choices based on what the speaker believes this specific partner will understand. In understanding, such adjustments could involve using what the addressee knows about this specific speaker to avoid any ambiguities, better recognize the speaker's intentions, and guide relevant inferences.

2. *Cultural/community/group-based adjustments*. In speaking, speakers make choices based on what they believe a member of a par-

ticular group (age, gender, ethnicity, club, neighborhood, family) can understand. For example, bilinguals mix words from both languages only to addressees with whom they share those languages and not to monolingual speakers. In understanding, addressees take the group context into account in interpreting specialized language (e.g., jargon, slang) and making relevant inferences.

3. *Generic partner adjustments.* Speakers make choices that the typical user of the language would find easiest to comprehend; addressees make assumptions for processing that would be warranted for any speaker of the language.[4]

4. *No adjustments (egocentric processing).* Processing is based on speakers' or addressees' own ease of production or comprehension. This may or may not benefit a partner, too.

Note that the adjustments that partners make may be based on information copresent to them (that the partners mutually know), or it can be based on information that one partner happens to know about the other (not in common ground).

Second, speakers and addressees might make these adjustments for any or all of the various aspects of spoken language use that psycholinguists have catalogued (see Bock, 1995; Levelt, 1989, for models of the processes involved in language production). Speakers might adjust word choice, pronunciation or other articulatory features (e.g., word duration), syntactic structure, sentence-level topic selection, or higher level discourse planning. For example, a New York native giving subway directions to a non-New Yorker could potentially select different words than she would when speaking with another New Yorker ("the number 2 train over there" rather than "the uptown IRT"). She might adjust her pronunciation for her addressee ("Forget about it" instead of "Fuggedaboudit"). She might clarify her syntax to maximize the chances of being understood amid the surrounding chaos, saying, "the train that you should take" rather than "the train you take"). She might organize her sentences to be as clear as possible to the uninitiated ("When you leave the station, go up the stairs to your left" rather than "Go to the northeast corner") and organize her entire discourse to be as helpful as possible ("This will be complicated—are you ready? First you take the number 2 train over there to 42nd Street, and transfer to the 1 or 9" rather than "Transfer to the local at Times Square").

Correspondingly, addressees might adjust their comprehension of speakers' intentions for any of these aspects of language use. For example, a father being addressed by his daughter must adapt to her nonadult pronunciation (*blanky* rather than *blanket*), word choice (*doggie* instead of *horse*), protosyntax (*up!* rather *please pick me up*), and difficult-to-interpret sequencing of information. Some of his adaptations are likely to be generic, others group level (what children his daughter's age are likely to say), and others quite individualized (what his daughter in particular says).

[4]Some generic adjustments might be context specific, as when speakers speak more loudly to addressees at a distance.

Third, if adaptations occur, their time course needs to be specified. Speakers and addressees who ultimately adapt to their partners may take their partner into account from the first moments of processing or they might adjust only later when they recognize a discrepancy. Or they might process several alternatives in parallel. How they do this may vary for different aspects of language use; that is, some aspects of language use (e.g., higher level discourse plans) may be more amenable to controlled adjustment than others (e.g., motor processing during articulation). Following Gerrig's (1986) distinction between *processes* and *products* of language use, we argue that both the processes and ultimate products of partner adaptation need to be elucidated. But the processes involved in partner adaptation are likely to unfold over longer stretches of time than most language researchers usually consider. Even the simplest models that conversationalists might create of each other would have to be built up as interaction progresses. To borrow from Gibbs (1994), the first few milliseconds of processing are important, but so are later moments of integration, interpretation, and appreciation.

These three variables—type of adaptation, aspect of language processing, and time course—define a complex potential space of partner adaptation. To further complicate matters, in actual conversation, additional factors are likely to come into play. Conversationalists' *knowledge* of differences in perspective, *ability* to see things from the partner's point of view, and *motivation* for doing so can affect when and how they adapt to each other.

Knowledge

Speakers and addressees can adapt to each other only to the extent that they are aware of differences in their perspectives and informational needs. To carry out group/community-based adaptation, they need to be aware of which groups their partners belong to and understand what pieces of knowledge come with membership in those groups. To carry out individualized adaptation, they need to know the individuals' characteristics. Sometimes the immediate physical environment supports the knowledge that conversational partners need in

order to adapt; for example, adapting to one's partner's point of view when describing a location is relatively straightforward if one can see one's partner (Schober, 1993, 1995). But keeping in mind the partner's group membership, conceptual preferences, conversational agenda, or world view is much less straightforward, as it relies less on direct or perceptual evidence and more on memory and inferences (Schober, 1998a).

In fact, the evidence is that people are far from perfect at estimating others' knowledge. Undergraduates who were asked to rate public figures for the extent to which the figures would be known to other undergraduates overestimated the identifiability of figures they themselves knew and underestimated the identifiability of figures they didn't themselves know (Fussell & Krauss, 1992). As Nickerson's (1999) review shows, this tendency for people's estimates to be biased in the direction of their own knowledge has been found in various knowledge domains (see e.g., Bromme & Nückles, 1998; Fussell & Krauss 1991, 1992; Nickerson, Baddeley, & Freeman, 1987). In some contexts, people can even be shown to be biased in assessing their own knowledge accurately, claiming to have known more in hindsight than they did earlier and generally overestimating what they know (for reviews, see e.g., Keren, 1991; Lichtenstein, Fischhoff, & Phillips, 1982).

To adjust appropriately, not only must speakers and addressees assess each other's knowledge, but they also must know what the appropriate adjustments are and when to deploy them. If they don't, they are liable to adjust ineffectively (see Horton & Gerrig, in press). The native speaker who speaks more loudly for the non-native addressee has appropriately assessed the partner's group membership and is carrying out a group-based adjustment, but she has inappropriately judged what the right adjustment might be (in fact, it may be the only adjustment she is capable of making). Similarly, an addressee who knows the speaker is a child may incorrectly assume that all single-word utterances are requests, although he has correctly assessed the speaker's group membership and lack of adult syntactic ability.

Ability

Even when people are fully knowledgeable about their conversational partner's perspective, they may vary in their momentary or chronic levels of ability to adapt to their partners, at least within particular domains. At the momentary level, conversationalists under greater cognitive load may have fewer attentional resources available to devote to the work that adapting to a partner requires (Flavell et al., 1968; Horton & Keysar, 1996). This is likely when discourse tasks are particularly difficult (Bard et al., 2000), raising the possibility that the nature of the discourse task could play a role in contradictory findings about partner adjustments.

At the chronic level, some people may be more egocentric than others—that is, they may lack general perspective-taking skills (Davis, 1983). Others may lack the expertise or ability within a particular domain of discussion that would enable them to take another's point of view. For example, in one study (Schober, 1998b), people with low spatial (mental rotation) ability were far less likely to accurately produce or comprehend descriptions of locations from their partner's point of view than people with high spatial ability.

Motivation

Even if people know their partner's perspective and are able to adapt to it, they may not be willing to, depending on their goals. People who are not getting along are, in fact, likely to *diverge* on various features of their speech as conversation proceeds (Bly, 1993; Danet, 1980; Giles, Coupland, & Coupland, 1991). According to Brown and Levinson (1987), adapting to a partner's perspective is usually more polite than requiring the partner to adapt because it imposes less on the partner. Speakers continually monitor each other to achieve the desired levels of politeness; if they don't intend to be polite, they adapt less. This is supported by findings like Graf's (described in Herrmann, 1988) that students describing the location of a plant's position in a room to an addressee represented by a puppet took their partner's perspective far more often when they were told that their addressee represented a professor than when it was supposed to be a fellow student. Presumably this kind of influence on adaptation is fairly high level and controlled.

Finally, we should note additional complexities in the notion of egocentrism. What may look like egocentric behavior could actually turn out to be the result of sophisticated consideration of one's partner (Schober, 1998a). For example, imagine that a speaker with low spatial ability is fully aware that his partner has high spatial ability. If he chooses to describe a location from his own perspective on the assumption that his partner can easily see things from his point of view, has he really behaved egocentrically? At some level, such a speaker has considered his own and his partner's needs and has made a judgment that reflects the balance between the two (or that minimizes the collaborative effort they expend together), which hardly seems like an egocentric judgment. Of course the speaker's behavior alone doesn't allow us to know whether this has happened (see Keysar, 1997), but the point is that just as what looks like a partner adjustment may not

really be one, what looks like egocentric behavior may not really be egocentric.

A related complication to the notion of egocentric processing is the necessity of distinguishing between truly not knowing or processing the partner's point of view and constraint-based processing. In constraint-based processing, listeners use whatever information is available to them at the moment, including, for example, information related to a speaker's goals that may serve to determine what a speaker means by an ambiguous referring expressions (see Hanna, 2001). If egocentric information is more accessible, it may win out. But this is no guarantee that the partner-based information wasn't being processed simultaneously (see Bard & Aylett, 2000; Hanna et al., 2002).

From these various theoretical considerations, it is clear that at any one moment of production or comprehension, people may be taking many sources of information into account at different stages of language processing. Various situational features could affect how much or in what ways a particular partner model is elaborated. The extent to which conversationalists pay attention to the relevant details of the conversational partner's knowledge is likely affected by both parties' conversational agendas, how much they like each other, the nature and the importance or difficulty of the task at hand, and various other factors.

THE ROLE OF THE PARTNER IN LANGUAGE PROCESSING: EMPIRICAL FINDINGS

With these theoretical distinctions and caveats in mind, we next lay out relevant empirical findings about whether, how, and when speakers and addressees adjust to each other in processing. These studies have varied speakers' and addressees' knowledge systematically to look at online effects of partners' knowledge on utterance production, utterance interpretation, and language use in conversation. The challenge these studies face is establishing the experimental control required to measure processing while maintaining the realistic interpersonal settings in which interactive spoken discourse occurs. The findings are inconclusive, with some studies suggesting that speakers or addressees do not take each other into account in the earliest moments of processing and others suggesting they do.

Effects of Addressee Knowledge on Speaking

An early study (Brown & Dell, 1987; Dell & Brown, 1991) focused on how speakers express typical and atypical information when the addressee has more or less ability to infer this information. The rationale

was that instruments that are easily inferrable (as *knife* is from *stab*) are often not mentioned explicitly because they represent predictable attributes in a schema, whereas atypical instruments that depart from the prototypical value of an attribute (as *ice pick* does in a stabbing) are tagged and highly available during utterance planning. From the speaker's perspective, this means that an atypical instrument is more likely to be mentioned explicitly than a typical instrument, as well as more likely to be mentioned in the same syntactic clause with the main action. The interesting question is whether the pattern of mention is affected by the needs of the addressee.

Brown and Dell had speakers retell very short stories to confederate addressees, to examine how the speakers chose to mention the instruments used to perform the stories' main actions. Storytellers saw a picture illustrating each story along with its instrument. In half of the pairs, the addressee saw a copy of the same picture; for the other half, the addressee had no picture, and the speaker was aware of this. The reasoning was that if storytellers simply did what was easiest for themselves, they should mention atypical instruments more often than typical ones regardless of whether their addressee could see a picture. If, on the other hand, storytellers adjusted to their addressees, they should be most likely to mention atypical instruments when their partners did not have pictures because there would be no way for the addressee to infer the instruments.

Brown and Dell concluded that storytellers' choices of whether and how to mention instruments were impervious to addressees' needs, at least in early utterance planning. That is, atypical instruments were mentioned more often than typical instruments as expected (and more often within the same clause with the target verb than typical instruments), but whether the addressee could see a picture did not seem to matter. The only reliable addressee effect was that storytellers mentioned instruments in separate clauses after the verb more often when the addressee could not see a picture than when the addressee could. Brown and Dell proposed that this represented a relatively late adjustment or repair on the part of the storytellers. They reported no evidence for any early adjustments to addressees.

This experiment has the merit that it involved an actual addressee who was introduced to the storyteller as another experimental participant, unlike in many studies of language production with no explicit audience at all. However, no detail was provided about the confederate's responses during the storytelling task; this is of concern because the latency and content of an addressee's feedback can affect what speakers say. For instance, in other experiments that used a storytelling task in which addressees' feedback was disrupted by having to do

a distracting secondary task, speakers told the stories in less detail (Bavelas, Coates, & Johnson, 2000; Pasupathi, Stallworth, & Murdoch, 1998). Thus, although the use of confederate addressees can be carefully staged so that subjects do not report suspicions about their partners, it is possible that confederates may not always provide the same kind of feedback that uninformed addressees do. This caveat applies to the Brown and Dell study, in which two confederates each heard the same stories dozens of times. Under normal circumstances, when addressees believe they already know what speakers are saying, they would probably deliver acknowledgments more quickly than when the information is new to them (see Lockridge & Brennan, 2002). If the confederate's backchannels were too prompt, the speakers may have concluded that their addressee understood the stories all too well regardless of whether they had visual copresence.

A second feature of Brown and Dell's study that may have affected the results is that it simulated physical copresence; speakers and addressees had separate displays and were not able to easily make eye contact and monitor each other's attention to the same display. A final caveat is that the findings of Brown and Dell's most relevant to our questions were null findings, which are difficult to interpret. One relevant comparison did approach significance: Speakers were marginally (p .09) less likely to explicitly mention either sort of instrument (typical or atypical) when addressees had pictures than when they did not. If this effect were reliable, it could be construed as a coarse adjustment to addressees' having pictures.

Lockridge and Brennan (2002) used Brown and Dell's method to look for addressee-centered adjustments with genuine (naive) addressees. Contrary to what Brown and Dell found, storytellers adjusted what they said according to whether addressees could see the pictures illustrating the stories. When addressees had no way to infer the instruments—that is, when they had no pictures and the instruments were atypical—speakers were about 10% more likely to mention the instruments in the same syntactic clause with the target action verb. This small but reliable effect suggests that speakers' choices in early utterance planning are not fully dictated by addressees' needs, but are not impervious to those needs either.

Other studies have examined the definiteness of referring expressions as signals by speakers about whether information is new or already known. In their studies, Clark and Wilkes-Gibbs (1986; Wilkes-Gibbs & Clark, 1992) reported that speakers were most likely to mark referring expressions as new (with an indefinite article) on the first reference to an object in conversation. Furthermore, the use of indefinite referring expressions was sensitive to addressees' participatory status in previous conversation (Wilkes-Gibbs & Clark, 1992). In Wilkes-Gibbs and Clark's study, directors and matchers did a card-matching task while observed by either bystanders (whom directors knew about but could not see) or side participants (who sat close by the director and visibly observed the task). When the side participants or bystanders were later paired with the directors to perform the same task, the directors tended to mark their initial referring expressions as new (with indefinite articles) with matchers who had been bystanders and as given (with definite articles) with matchers who had been side participants. This is a clear adjustment to addressees' needs because the objects marked as new were new only to the addressees (in immediately previous trials with different matchers, the speakers had used definite references). Consistent evidence that speakers adjust the definiteness of referring expressions to a partner's needs comes from Lockridge and Brennan's (2002) experiment: Speakers were more likely to mark their references to instruments as definite when addressees had pictures or when the instruments were typical, and they were more likely to mark references as indefinite when addressees had no pictures and when instruments were atypical.

It is reasonable to ask whether adjusting the definiteness of a referring expression to an addressee's knowledge is done early or late in planning. Wilkes-Gibbs and Clark's (1992) and Lockridge and Brennan's (2002) studies did not control the timing by which speakers made their choices. Both studies examined *first* spontaneous mentions, which does not of course rule out that speakers may have hesitated prior to a noun phrase in order to complete a monitoring and adjustment process of the sort suggested by Horton and Keysar (1996). Yet in Lockridge and Brennan's study, the apparent adjustment to an addressee having a picture was made just as often when the first mention of the instrument was in the same clause as the main action as when the first mention was later in the utterance. On the assumption that entities available earlier are lexicalized earlier (Bock, 1995; Levelt, 1989), this suggests that information about addressees' knowledge was available relatively early on.

In a reanalysis of corpus data presented in Bard et al. (2000), Bard and Aylett (2000) coded referring expressions for definiteness using a hierarchy inspired by Ariel (1990) and Gundel, Hedberg, and Zacharski (1993). When a speaker produced a referring expression for the second time but directed it at a new addressee, the expression was no more likely to be marked as given than it was on being directed to the first addressee. This (null) result is consistent with Clark and Wilkes-Gibbs (1986; Wilkes-Gibbs & Clark, 1992) and Lockridge and Brennan (2002), further supporting a conclusion that the marking of

given and new status using definite and indefinite articles is sensitive to the cognitive state of the addressee, rather than the cognitive state of the speaker alone.

With respect to a different sort of syntactic choice, a series of carefully designed studies by Ferreira and Dell (2000) examined whether speakers make optional syntactic choices in a way that could actually reduce ambiguity for listeners ("ambiguity-sensitive sentence production") or whether these choices are based on what is easiest for speakers ("availability-based sentence production"). In these studies, speakers were prompted to reproduce sentences they had heard earlier in which optional complementizers were either present or omitted (e.g., *that* in *The coach knew ____ you . . .*). When these complementizers are absent, such sentences can be temporarily ambiguous for addressees because at least for a moment they allow either a direct object interpretation of the following pronoun (e.g., *The coach knew you when you were younger*) or a subject interpretation (e.g., *The coach knew you were the best tennis player in town*). This ambiguity is eliminated when the pronoun has clear case marking (e.g., *The coach knew me . . .* vs. *The coach knew I . . .*). If utterance design is availability based, speakers should produce a complementizer to fill the time when the following word is not yet available enough to be articulated, but not when the following word is already activated. If utterance design is ambiguity-sensitive, speakers should be more likely to produce a complementizer when the following word is ambiguous as to case (*you* rather than *I*): without the *that*, the incremental utterance (e.g., *The coach knew you . . .*) would be temporarily ambiguous to the addressee. Availability was manipulated by which words were used as recall cues. The same paradigm was also used to examine speakers' production of optionally reduced or full relative clauses.

In five of six experiments, Ferreira and Dell had speakers talk into a microphone while facing a screen with a visual display of a deadline within which they had to finish producing the sentence; the experimenter sat next to them. In the sixth experiment, half of the speakers did this while the other half addressed live partners who were given the task of rating the sentences for clarity. The results of all six experiments supported the availability-based (speaker-centered) hypothesis: Speakers were no more likely to use optional function words in ambiguous than unambiguous sentences. However, in the sixth experiment, speakers *did* produce slightly more (7%) complementizers overall when they spoke to naive addressees (significant by-items but not by-subjects) than when they produced their utterances only for the experimenter. If speakers indeed use more optional function words overall with overt addressees, this could be construed as an adjustment to

their needs, albeit not a sensitive one because the use of complementizers was still unrelated to whether the sentences were temporarily syntactically ambiguous. Of course the question remains as to whether the sentence recall task performed with a mute addressee who would later be given a rating task is sufficiently like what speakers do spontaneously with interacting addressees (as the authors themselves pointed out). It is also possible that sentences can be disambiguated using prosody, which was not measured in these experiments.

Next we turn to an innovative experiment by Horton and Keysar (1996) explicitly aimed at teasing apart speakers' and addressees' distinctive (or privileged) knowledge and examining the time course by which knowledge about a partner's knowledge is used. In that experiment, subjects referred to a target object in the context of a background object that they were told an addressee either could or could not see. The addressee was a confederate seated behind a barrier. Speakers appeared to take addressees' knowledge into account only when they were allowed to form referring expressions without time pressure. When placed under time pressure, speakers were just as likely to produce the adjective contrasting the target object (e.g., *big* in "the big square") from its background object (a smaller square) when the addressee could see the background object as when the addressee could not. These findings were interpreted as supporting a fast, automatic stage of egocentric processing in utterance planning and a slower stage of monitoring and adjusting the utterance for appropriateness. Presumably these two processes could happen in parallel, although Horton and Keysar focused on serial processes of egocentric processing followed by monitoring and adjustment.

This experiment has generated some controversy (see Polichak & Gerrig, 1998; Keysar & Horton, 1998) for several reasons. One critique is that the task afforded no physical copresence between the partners. Speakers in the privileged knowledge condition were required to keep in mind that their partner could see one part of the display (the moving part) but not the other (see Polichak & Gerrig, 1998). It would not be surprising if having to keep track of which information was and was not known to a partner, in the absence of perceptual cues to visual copresence, was disrupted by pressure to speak quickly. In addition, as the target object moved off the subject's screen onto the confederate addressee's screen, it sometimes changed shape (the confederate's task was to identify whether the object was the same as or different from the one the speaker had described), which also violates the physical properties of most shared visual environments. Speakers may be better able to represent and adjust to an addressee's perspective when there is actual or potential physical copresence between them.

In contrast to Horton and Keysar's findings with adults, Nadig and Sedivy (2000) found that even 5- to 6-year-old children (typically assumed to be more egocentric than adults) take into account the common ground they share with addressees. Children had to refer to a target object within a field of three background objects that included (a) a competitor object similar to the target (e.g., a big and a small cup) that was also visible to the addressee (and so was in common ground), (b) a competitor similar to the target but invisible to the addressee (in privileged ground), and (c) an unrelated object in privileged ground. These conditions provided a similarity control not found in the Horton and Keysar experiment. Children provided informative adjectives to disambiguate the similar targets more often in (a), when common ground included a similar competitor, than in (b), when the similar competitor was in privileged ground, as well as in (c), when there was no similar competitor.

Now we move from the level of word and syntax choices to variations in articulation. A series of studies by Bard et al. (2000; Bard & Aylett, 2000) presented evidence that articulation is an egocentric process, in which intelligibility is adjusted based not on what addressees have heard previously, but on what speakers themselves have heard. They compared intelligibility of repeated words by the same speaker talking to two different addressees in succession, finding that speakers shortened repeated words on re-referring, even though their second addressee was hearing them for the first time. This suggests that speakers do not actually take their addressees' needs into account when articulating given and new information, despite widespread assumptions to the contrary (e.g., Nooteboom, 1991; Samuel & Troicki, 1998). With the same corpus, Bard and Aylett (2000) found that, although speakers adjusted the definiteness of their descriptions to new addressees, they did not adjust their articulation, which Bard and Aylett argued supports the dual-process model. Unfortunately, the corpus didn't enable an important control: repeated mention to the same partner.[5] That is, intelligibility degraded slightly with a new partner, but it might have degraded even more with an old partner. Recall that McAllister et al. (1994) found that repeated words had shorter durations in dialogues (with an addressee present) but not in monologues, suggesting that partner adjustments may be possible at the articulatory stage after all.

[5]Particularly interesting is the Bard et al. (2000) finding that when the first mention of a word is by one partner and the second mention is by the other, the second mention is shortened as if it had been produced by the first partner. This suggests that both the speaker's own and her partner's utterances affect the speaker's representations of given and new information.

A study by Schafer, Speer, Warren, and White (2001) examined the use of prosody in a task-oriented discourse context, finding that speakers indeed used prosody to disambiguate prepositional phrases with high versus low attachment, but that the use of this prosodic contrast was independent of whether the prepositional phrase would actually have been ambiguous to addressees in the syntactic context.

In sum, the evidence about whether and how speakers adjust to addressees is mixed, due in part to the difficulty of teasing apart the addressee's knowledge from the speaker's, to variations in how copresent the partners are and how realistic the tasks are, to other confounds and missing controls, and to the difficulty of interpreting null effects. It seems clear that some types of adjustments are more feasible than others. Taken at face value, the findings suggest that articulation and some syntactic choices are more automatic and less influenced by addressees' needs than are other syntactic choices, referring expressions, and their marking as definite or indefinite.

Effects of Speaker's Knowledge on Addressees' Interpretation

An extensive program of experiments by Keysar and colleagues has tested whether interpretation is restricted to only that information presumed to be in common ground or whether privileged knowledge (information known only to the addressee) interferes. In the first Keysar et al. (1998) study, subjects who acted as addressees were given information on a computer screen (e.g., "Rachel delivered the sofa") with which to answer yes or no to confederate speakers' questions (e.g., either "Did Rachel deliver the sofa?" or "Did she deliver the sofa?"). The questions were prerecorded (but were assumed by subjects to be live) and delivered over an intercom. Subjects also had a second task in the guise of increasing their memory load; they had to remember a second statement that was presented right after the relevant information (e.g., "Marla delivered a cake"). They were told this information was irrelevant to the task and unknown by the speaker (making it privileged to the subject). The key comparison was whether subjects would be slower to answer questions containing pronouns that were ambiguous between referents in the shared and privileged sentences. Indeed, the privileged information (e.g., Marla) interfered with the information in common ground (e.g., Rachel) in the pronoun/same gender condition, yielding longer latencies and increased errors.

In the second Keysar et al. (1998) study, a live confederate served as speaker and the addressee's gaze was tracked, replicating the interference of privileged knowledge. These results were used to rule out a restricted search hypothesis, concluding that addressees do not restrict

their search for referents to only that information presumed to be in common ground.

In both of these studies, however, the privileged information that competed with the information in common ground was not only ambiguous, but strongly favored by recency. It would have been remarkable if subjects had been able to suppress this salient information entirely; interference between similar items is pervasive in the human memory system. The question remains: If both privileged information and common ground were to start on a level playing field, would common ground tend to be consulted first?

In a study with a confederate speaker speaking from a script to naive addressees (Keysar, Barr, Balin, & Brauner, 2000), a vertical display of see-through cubbyholes held objects that were either visible to both speaker and addressee (common ground condition) along with objects occluded from the speaker (addressees' privileged ground condition). The goal was to test a version of the *optimal design* principle—that addressees would search only common ground when the speaker made a reference that was ambiguous between an object in common ground and another in privileged ground. Addressees' first looks were not limited to what was in common ground, but went to objects in privileged ground just as often. The authors concluded that initial processing is egocentric.

However, in this study the critical instructions were biased in favor of the objects in privileged ground over those in common ground. That is, with "put the smaller candle. . . ," there was one large candle, a smaller candle in common ground (the speaker's "intended" referent), and an even smaller candle in privileged ground (for discussion, see Hanna et al., 2002). The Hanna et al. (2002) study removed this bias so that the privileged object was not the most typical referent for the critical referring expression. When this bias was removed, first looks were made significantly more often (although not solely) to the object in common ground than to the one in privileged ground. This indicates that common ground does not operate only as a late filter, but can be used in the earliest moments of processing.

Apparently small children can use common ground in comprehension as well. To return to Nadig and Sedivy's (2000) eye-tracking study, 5- to 6-year-olds were instructed to pick up a target object (e.g., the cup) when there was (a) a similar competitor also visible to the speaker that rendered the expression ambiguous (e.g., a second cup), (b) a similar competitor that was visible only to the child and thus in privileged ground, and (c) an unrelated competitor. The children's eye movements showed significant numbers of looks to the similar competitor in (a) (where the reference to the cup was infelicitous), but no interference from the similar competitor that was invisible to the speaker in (b). This evidence is consistent with the interpretation that common ground helps constrain the interpretation of referring expressions from the earliest moments of processing; the fact that it can do so in children this young is remarkable.

Interpreting the Findings About Speakers' and Addressees' Apparent Adjustments and Failures to Adjust

These empirical findings leave us with the proverbial problem of deciding whether the glass is half full or half empty. Should speakers and addressees be characterized as adapting to each other well or even adequately? Or should they be considered mired in egocentricity? The trouble is that data from the same studies can be given alternate spins, either focusing on people's successful adaptations or on their mistakes.

Consider the data from Clark and Schaefer's (1987) study, in which pairs of undergraduate friends were given the task of describing campus landmarks for each other in ways that would conceal their identity from overhearers. They managed to develop successful private keys about 50% of the time. But the rest of the time they leaked far more than they realized; overhearers could guess the referents fairly accurately based on descriptions the pairs thought would be private (like "This is where someone wanted to put my teddy bear" to refer to a fountain). One spin on these results is that students were moderately successful at concealing information from others, but the alternate spin is that they made a great number of errors.

Or consider Lockridge and Brennan's (2002) finding that speakers adapt to their addressees' needs in instrument mention. The raw percentage of adaptation, although statistically reliable, was actually fairly small; speakers were 10% more likely to mention atypical instruments early in syntactic planning when addressees lacked pictures. Again, one spin is that speakers adapted to their listeners; another is that the adaptation was relatively modest.

The truth of the matter is obviously that people can fail to adapt to each other. We have all been in situations where speakers have seemed oblivious to the knowledge or informational needs of their audiences; think of the jargon-spewing lecturer who assumes too much (or the condescending lecturer who assumes too little) or the person in the cinema who somehow can't seem to whisper. And we have all been in situations when addressees have failed to consider the source as they interpreted speakers' utterances, finding insult where none was intended, misinterpreting small talk as serious talk, or assuming they

understand what is beyond their level of expertise. Experimental findings on referential communication from the beginning have demonstrated that people don't adapt to each other in every way they possible could. Glucksberg et al. (1966) found that 4-year-old speakers in a dyadic communication task produced messages that were unintelligible to others, although the messages were intelligible to the speakers. More recently, Russell and Schober (1999) found that speakers who were not informed that their partners' task goals differed from theirs persisted in assuming that their partners shared their goals, ignoring mounting evidence in the discourse that there were problems (see also Schober, 1998b).

In fact, although people can form effective conceptual pacts that allow them to refer efficiently in the task at hand, there is no guarantee that successful referring entails truly aligned conceptualizations. Sometimes people in conversation do not detect important mismatches. In studies of how respondents in standardized surveys comprehend seemingly banal questions about facts and behaviors (Conrad & Schober, 2000; Schober & Conrad, 1997), respondents interpreted ordinary words like *furniture*, *bedroom*, and *job* quite differently than the survey designers intended a substantial portion of the time (see also Belson, 1981, 1986). It almost never occurred to survey respondents that their interpretations could possibly be different than the interviewers', even with extensive instructions to that effect. Survey respondents seemed to be relying on a *presumption of interpretability* (Clark & Schober, 1991)—that if their own interpretation was insufficient, the interviewer would have somehow made this clear.

The moral seems to be that people's apparent success at a communicative task doesn't in and of itself provide evidence about how—or if—they have taken each other into account. For example, in one study, survey respondents who were asked "Have you smoked at least 100 cigarettes in your entire life?" answered "yes" or "no" without hesitation; they seemed to be communicating successfully. Although they never suspected that their notions of what counted as smoking (finishing cigarettes? taking a puff?) or cigarettes (tobacco? marijuana? cloves? cigars?) might differ from other people's, 10% of them changed their answer to the question when provided with a uniform definition later (Suessbrick, Schober, & Conrad, 2000). Similarly, people's communicative failure doesn't show that they *haven't* taken each other into account. For instance, when speakers in phone conversations suddenly discover that they don't mean the same person by a proper name, it often isn't because they have failed to assess their partner's knowledge appropriately, but rather that they have failed to produce an appropriate cue to that knowledge (Horton & Gerrig, 2002).

CONCLUSIONS

Much remains to be discovered about when and how speakers and addressees adapt to each other. In particular, we know little about the mental representations involved in partner adaptation. When there is a model of a partner's characteristics or beliefs, we don't know how elaborate or coherent it is, nor what role it plays in cognition in general. We don't know exactly how detailed the representation is, how specific it is to an individual, or how often it is updated.

What is striking from the literature on how and when speakers and addressees take each other into account is how much more complex the issues are than they at first seem. Given the complexities, it seems naive to imagine that we can determine, across discourse contexts, and for a particular aspect of language use (say word selection in production), what sorts of adaptations always occur. Presumably the adaptations vary according to conversationalists' (chronic or momentary) attentional capacities, discourse goals, interest in taking their partner's perspective, and so on. It is tempting to set the research agenda in traditional terms: to examine the limits on what speakers *could* do. That is, if we could discover that one aspect of processing is entirely modular and *can't* allow initial partner adaptation, this would constrain our theory of just how much speakers can adapt. But such a research agenda is problematic, not least because it leads to a search for null findings (and it is logically impossible to show that something isn't possible). A traditional research agenda would also require researchers to generalize across too many unspecified variables, as any one experimental setting can only look at one small corner of the space of possibilities.

The evidence so far suggests that adaptation doesn't seem to be an all-or-nothing phenomenon at any level; people can be shown to adapt under some circumstances and not to adapt under others at virtually every level of language use—from higher discourse-level functions to articulation. Thus, we propose, the more fruitful research agenda is to explore the factors that affect conversationalists' adaptations in particular circumstances—the sorts of tasks, individual ability differences, discourse goals, and so on that affect when and how partners can adapt to each other.

Systematic research always requires trade-offs between naturalism and experimental control. But the methodological issues in studying interactive spoken discourse are particularly thorny because the challenge is to model people's individual mental processes while they are simultaneously involved in the collective activity of conversation. The trade-offs that experimental psychologists often make—for example,

having participants listen to prerecorded disembodied utterances so as to control the stimulus—simply won't do for studying processing during interactive discourse because the prerecorded utterances aren't contingent on the participants' actions. An interacting scripted confederate may be effective for some purposes, but only if the confederate's conversational feedback is as contingent and plausible as a naive partner's would be. The tasks prescribed by an experimenter are necessary for control, but unless these tasks contact what people do naturally during communication they may not generalize. Obviously intrusive online tasks like lexical decision are unsuitable for studying interactive discourse. The use of relatively nonintrusive head-mounted eye-tracking devices is extremely promising, but the eye-tracking method brings its own methodological challenges. Ultimately, a complete theory of when and whether partners adapt to each other in conversation requires researchers to use a wide variety of tasks that involve different kinds of goals and affordances.

An even greater challenge is to be clear about theoretical distinctions that haven't always been made. Researchers must distinguish individual and mutual knowledge. They must be clear about what sorts of common ground and individual knowledge are involved in any particular interaction. Perceptual copresence, for example, potentially provides several kinds of information, any of which might make a difference in processing; participants can gain knowledge about their partners by seeing whether they look confused, what they appear to be looking at or pointing to, what progress they have made in a task, and so on. Experiments that set up scripted situations in which conversationalists can both see some of the same physical objects while individually seeing privileged competitor objects need to be clear about the relative salience of those objects (how typical they are, how relatively well they fit the referring expressions, how recently they were mentioned, and how they were previously mentioned). Studies examining situations where speakers have common ground based on linguistic copresence or community comembership must be clear about exactly what sorts of inferences participants must make, taking care that the experimental demands aren't too different from real-life attentional demands.

Finally, experimenters need to recognize that laboratory settings themselves are social situations with their own logic and intentional structure (for explicit discussion, see Schwarz, 1996, 1998). To the extent that experimental participants recognize that there is an extra agenda-laden audience—the experimenter—for anything they do in the lab, the seemingly simple two-participant experiment may actually become a complex multiagent situation in which any utterance may involve more than one level of intentionality. How people adapt to their partners in such circumstances may or may not reflect what they would do unobserved. The challenge for studies of interactive spoken discourse, then, is to make appropriate design trade-offs and really understand the language game in which experimental participants find themselves.

ACKNOWLEDGMENTS

We are grateful to many colleagues for helpful comments and debate concerning this chapter. This material is based on work supported by the National Science Foundation under Grants No. 0081550, 0082602, 9980013, 9711974, and 9730140. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Ariel, M. (1990). *Accessing noun-phrase antecedents*. New York: Routledge.

Austin, J. L. (1962). *How to do things with words* (2nd ed.). Cambridge, MA: Harvard University Press.

Bard, E. A., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Bard, E. A., & Aylett, M. P. (2000). Accessibility, duration, and modeling the listener in spoken dialogue. *Götelog 2000, Fourth Workshop on the Semantics and Pragmatics of Dialogue*. Götalog, Sweden: Gothenburg University.

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941–952.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*, 145–204.

Belson, W. A. (1981). *The design and understanding of survey questions*. Aldershot: Gower.

Belson, W. A. (1986). *Validity in survey research*. Aldershot: Gower.

Blakar, R. (1984). *Communication: A social perspective on clinical issues*. Oslo, Norway: Universitetsforlaget.

Bly, B. (1993). *Uncooperative language and the negotiation of meaning*. Unpublished doctoral dissertation, Stanford University.

Bock, J. K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181–216). San Diego: Academic Press.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes, 23*, 119–147.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in spontaneous speech: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*, 123–149.

Boyle, E., Anderson, A., & Newlands, A. (1994). The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language and Speech, 37*, 1–20.

Brennan, S. E. (1988). The multimedia articulation of answers in a natural language database query system. *Proceedings, Second Conference on Applied Natural Language Processing* (pp. 1–8). Austin, TX: Association of Computational Linguistics.

Brennan, S. E. (1990). *Seeking and providing evidence for mutual understanding*. Unpublished doctoral dissertation, Stanford University.

Brennan, S. E. (1995). Centering attention in discourse. *Language & Cognitive Processes, 10*, 137–167.

Brennan, S. E. (2000). Processes that shape conversation and their implications for computational linguistics. *Proceedings, 38th annual meeting of the Association for Computational Linguistics*. Hong Kong: ACL.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1482–1493.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.

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*, 383–398.

Bromme, R., & Nückles, M. (1998). Perspective-taking between medical doctors and nurses: A study on multiple representations of different experts with common tasks. In M. Van Someren, P. Reimann, E. Boshuizen, & T. DeJong (Eds.), *Learning with multiple representations* (pp. 175–196). Oxford: Elsevier.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge, England: Cambridge University Press.

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*, 441–472.

Carroll, J. M. (1980). Naming and describing in social communication. *Language and Speech, 23*, 309–322.

Clark, H. H. (1992). *Arenas of language use*. Chicago, IL: University of Chicago Press.

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

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

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes, 23*, 567–598.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 313–330). Hillsdale, NJ: Lawrence Erlbaum Associates.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*, 332–373.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.

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. (1989). Contributing to discourse. *Cognitive Science, 13*, 259–294.

Clark, H. H., & Schober, M. F. (1991). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 15–48). New York: Russell Sage Foundation.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22*, 245–258.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201–242.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*, 1–28.

Danet, B. (1980). "Baby" or "fetus"?: Language and the construction of reality in a manslaughter trial. *Semiotica, 32*, 187–219.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113–126.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener." In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deutsch, W., & Pechmann, T. (1982). Social interaction and development of definite descriptions. *Cognition, 11*, 159–184.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*, 104–113.

Ferreira, V. S., & Dell, G. S. (2000). The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*, 296–340.

Flavell, J. H., Botkin, P. T., Fry, C. L., Jr., Wright, J. W., & Jarvis, P. E. (1968). *The development of role-taking and communication skills in children*. New York: Wiley.

Ford, W., & Olson, D. (1975). The elaboration of the noun phrase in children's descriptions of objects. *Journal of Experimental Child Psychology, 19*, 371–382.

Fowler, C., & Housom, J. (1987). Talkers' signalling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*, 489–504.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes, 27*, 35–53.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition, 62*, 151–167.

Fussell, S. E., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology, 19*, 509–525.

Fussell, S. E., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge. *European Journal of Social Psychology, 21*, 445–454.

Fussell, S. E., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*, 378–391.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Garrod, S., & Doherty, G. (1994). Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition, 53*, 181–215.

Gerrig, R. (1986). Process and products of lexical access. *Language and Cognitive Processes, 3*, 187–195.

Gerrig, R. (1994). *Experiencing narrative worlds*. New Haven, CT: Yale University Press.

Gibbs, R. W. (1994). *The poetics of mind*. Cambridge, England: Cambridge University Press.

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation*. Cambridge, England: Cambridge University Press.

Glucksberg, S., Krauss, R., & Weisberg, R. (1966). Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology, 3*, 333–342.

Goffman, E. (1976). Replies and responses. *Language in Society, 5*, 257–313.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Graesser, A. C., Person, N. K., Harter, D., & The Tutoring Research Group. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12*, 257–279.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics: Vol. 3. Speech acts* (pp. 225–242). New York: Seminar.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(2), 131–149.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*, 274–307.

Hanna, J. E. (2001). *The effects of linguistic form, common ground, and perspective on domains of referential interpretation*. Unpublished doctoral dissertation, University of Rochester.

Hanna, J. E., Tanenhaus, M. K., Trueswell, J. C., & Novick, J. M. (2002). *The effects of linguistic form and common ground on domains of referential interpretation*. Unpublished manuscript.

Herrmann, T. (1988). *Partnerbezogene Objektlokalisation—ein neues sprachpsychologisches Forschungsthema* [Partner-oriented localization of objects—a new psycholinguistic research topic]. (Bericht Nr. 25). Mannheim: University of Mannheim, Forschergruppe "Sprechen und Sprachverstehen im sozialen Kontext."

Horton, W. S., & Gerrig, R. J. (in press). Speakers' experiences and audience design: Knowing *when* and knowing *how* to adjust utterances to addressees. *Journal of Memory and Language*.

Horton, W. S., & Gerrig, R. J. (2002). *Common ground and memory processes in language production*. Manuscript under review.

Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.

Hupet, M., & Chantraine, Y. (1992). Changes in repeated reference: Collaboration or repetition effects? *Journal of Psycholinguistic Research, 21*, 485–496.

Hupet, M., Seron, X., & Chantraine, Y. (1991). The effects of the codability and discriminability of the referents on the collaborative referring procedure. *British Journal of Psychology, 82*, 449–462.

Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*(1), 26–37.

Jefferson, G. (1974). Error correction as an interactional resource. *Language in Society, 2*, 181–199.

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a standard maximum silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166–196). Clevedon: Multilingual Matters.

Karsenty, L. (1999). Cooperative work and shared visual context: An empirical study of comprehension problems in side-by-side and remote help dialogues. *Human-Computer Interaction, 14*, 283–316.

Keren, G. B. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217–273.

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

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

Keysar, B., Barr, D. J., Balin, J. A., & 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., & 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., & Horton, W. S. (1998). Speaking with common ground: From principles to processes in pragmatics: A reply to Polichak and Gerrig. *Cognition, 66*, 191–198.

Kingsbury, D. (1968). *Manipulating the amount of information obtained from a person giving directions*. Unpublished honor's thesis, Department of Social Relations, Harvard University.

Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology, 6*, 81–98.

Krauss, R. M., & Fussell, S. R. (1996). Social psychological models of interpersonal communication. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 655–701). New York: Guilford.

Krauss, R., & Glucksberg, S. (1969). The development of communication: Competence as a function of age. *Child Development, 40*, 255–266.

Krauss, R., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction. *Psychonomic Science, 1*, 113–114.

Krauss, R., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343–346.

Kraut, R., Lewis, S., & Swezey, L. (1982). Listener responsiveness and the coordination of conversation. *Journal of Personality and Social Psychology, 43*, 718–731.

Kraut, R. E., Miller, M. D., & Siegel, J. (1996). Collaboration in performance of physical tasks: Effects on outcomes and communication. *Proceedings, ACM Conference on Computer Supported Cooperative Work*. Cambridge, MA: ACM.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, England: Cambridge University Press.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*, 172–187.

Lloyd, P. (1991). Strategies used to communicate route directions by telephone: A comparison of the performance of 7-year-olds, 10-year-olds, and adults. *Journal of Child Language, 18*, 175–189.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review, 9*, 550–557.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language and Social Psychology, 7*, 181–191.

McAllister, J., Potts, A., Mason, K., & Marchant, G. (1994). Word duration in monologue and dialogue speech. *Language and Speech, 37*, 393–405.

McGregor, G. (1986). Listening outside the participation framework. In G. McGregor & R. S. White (Eds.). *The art of listening* (pp. 55–72). London: Croom Helm.

Moerman, M. (1987). Finding life in dry dust. In M. Moermann (Ed.), *Talking culture: Ethnography and conversation analysis* (pp. 19–30). Philadelphia: University of Pennsylvania Press.

Nadig, A., & Sedivy, J. (2000). Finding common ground in children's referential communication. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the 22nd annual conference of the Cognitive Science Society* (p. 1046). Mahwah, NJ: Lawrence Erlbaum Associates.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737–759.

Nickerson, R. S., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what others know influenced by what they themselves know? *Acta Psychologica, 64*, 245–259.

Nooteboom, S. G. (1991). Perceptual goals of speech production. *Actes du XIIème Congres International de Sciences Phonetiques, Vol. 8/19–24/91* (pp. 107–110). Aixes-en-Provence, France.

Olson, D. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257–273.

Pasupathi, M., Stallworth, L. M., & Murdoch, K. (1998). How what we tell becomes what we know: Listener effects on speakers' long-term memory for events. *Discourse Processes, 26*, 1–25.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89–110.

Poesio, M., & Vieira, R. (1998). A corpus-based investigation of definite description use. *Computational Linguistics, 24*, 183–216.

Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition, 66*, 183–189.

Reisman, K. (1974). Contrapuntal conversations in an Antiguan village. In R. Bauman & J. Sherzer (Eds.), *Explorations in the ethnography of speaking* (pp. 100–124). Cambridge, England: Cambridge University Press.

Rommetveit, R. (1974). *On message structure: A framework for the study of language and communication*. New York: Wiley.

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes, 27*, 1–33.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696–735.

Samuel, A. G., & Troicki, M. (1998). Articulation quality is inversely related to redundancy when children or adults have verbal control. *Journal of Memory and Language, 39*, 175–194.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2001). Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversational task. In *Proceedings, the 14th annual CUNY conference on Human Sentence Processing* (p. 89). Philadelphia, PA.

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, DC: 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. 266–296). Cambridge, England: Cambridge University Press.

Schegloff, E. A. (1987). Some sources of misunderstanding in talk-in-interaction. *Linguistics, 25*, 201–218.

Schegloff, E. A. (1988). Discourse as an interactional achievement: II. An exercise in conversation analysis. In D. Tannen (Ed.), *Linguistics in context: Connecting observation and understanding* (pp. 135–158). Norwood, NJ: Ablex.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 7*, 289–327.

Schiffrin, D. (1984). Jewish argument as sociability. *Language in Society, 13*, 311–335.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1–24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20*, 219–247.

Schober, M. F. (1998a). Different kinds of conversational perspective-taking. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive psychological approaches to interpersonal communication* (pp. 145–174). Mahwah, NJ: Lawrence Erlbaum Associates.

Schober, M. F. (1998b). *How partners with high and low spatial ability choose perspectives in conversation*. Paper presented at the 39th annual meeting of the Psychonomic Society, Dallas, TX.

Schober, M. F. (1999). Making sense of questions: An interactional approach. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangue (Eds.), *Cognition and survey research* (pp. 77–93). New York: Wiley.

Schober, M. F., & Carstensen, L. L. (2002). *Do age and long-term relationship matter in conversations about unfamiliar things?* Manuscript under revision.

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

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*, 576–602.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schwarz, N. (1998). Communication in standardized research situations: A Gricean perspective. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 39–66). Mahwah, NJ: Lawrence Erlbaum Associates.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, England: Cambridge University Press.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development, 38*(5), 1–38.

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

Stellmann, P., & Brennan, S. E. (1993). Flexible perspective-setting in conversation. *Abstracts of the Psychonomic Society, 34th Annual Meeting* (p. 20). Washington, DC.

Suessbrick, A. L., Schober, M. F., & Conrad, F. G. (2000). Different respondents interpret ordinary questions quite differently. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

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

Tannen, D. (1984). *Conversational style: Analyzing talk among friends*. Norwood, NJ: Ablex.

Tannen, D. (1994). *Gender and discourse*. New York: Oxford University Press.

Todorov, T. (1984). *Mikhail Bakhtin: The dialogical principle*. Minneapolis, MN: University of Minnesota Press.

Vygotsky, L. S. (1962). *Thought and language* (L. Mateka & I. R. Titunik, Trans.). Cambridge, MA: Harvard University Press.

Whittaker, S. J., Brennan, S. E., & Clark, H. H. (1991). Coordinating activity: An analysis of interaction in computer-supported cooperative work. *Proceedings, CHI '91, Human Factors in Computing Systems* (pp. 361–367). New Orleans, LA: Addison-Wesley.

Wilkes-Gibbs, D. (1986). *Individual goals and collaborative actions*. Unpublished doctoral dissertation, Stanford University.

Wilkes-Gibbs, D. (1995). Coherence in collaboration: Some examples from conversation. In M. A. Gernsbacher & T. Givón (Eds.), *Coherence in spontaneous text (Typological Studies in Language, 31)* (pp. 239–267). Amsterdam: John Benjamins Publishing Company.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183–194.

Yule, G. (1997). *Referential communication tasks*. Mahwah, NJ: Lawrence Erlbaum Associates.

Handbook of Discourse Processes

Edited by

Arthur C. Graesser
University of Memphis

Morton Ann Gernsbacher
University of Wisconsin, Madison

Susan R. Goldman
University of Illinois, Chicago

LEA
2003
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey
London