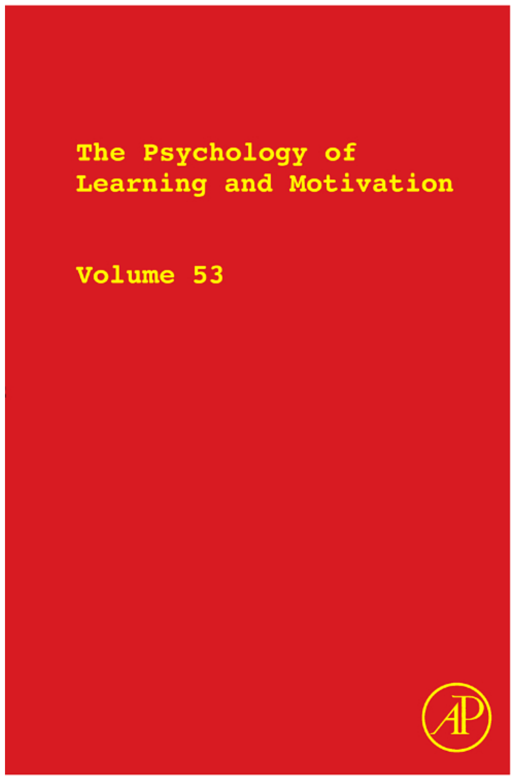

From: Susan E. Brennan, Alexia Galati, and Anna K. Kuhlen: Two Minds, One Dialog: Coordinating Speaking and Understanding
In Brian H. Ross editor:
The Psychology of Learning and Motivation, Vol. 53,
Burlington: Academic Press, 2010, pp.301-344.
ISBN: 978-0-12-380906-3

CHAPTER EIGHT

Two Minds, One Dialog: Coordinating Speaking and Understanding

Susan E. Brennan, Alexia Galati, *and* Anna K. Kuhlen

Contents

Abstract

In this chapter, we consider communication as a joint activity in which two or more interlocutors share or synchronize aspects of their private mental states and act together in the world. We summarize key experimental evidence from our own and others' research on how speakers and addressees take one another into account while they are processing language. Under some circumstances, production and comprehension are adjusted to a partner's perspective or characteristics in the early moments of processing, in a flexible and

Psychology of Learning and Motivation, Volume 53
ISSN 0079-7421, DOI: 10.1016/S0079-7421(10)53008-1

probabilistic fashion. We advocate studying the coordination and integration of cognitive products and processes both *between* and *within* the minds of interlocutors. We then discuss recent evidence from electrophysiology and imaging studies (relevant to Theory of Mind and to mirroring) that has begun to illuminate brain networks that underlie the coordination of joint and individual processing during communication.

1. Introduction: The Joint Nature of Language Processing

The scientific study of language has been shaped by the assumption that the human language faculty evolved for thinking rather than for communicating (e.g., Chomsky, 1965, 1980). This "language-as-product" tradition takes language itself as the object of study, focusing on grammatical knowledge and the core processes for recovering linguistic structure from sentences. This common focus has given generations of psycholinguists and other cognitive scientists license to concentrate on the study of the linguistic representation and processing in the mind and brain of a lone (and largely generic) native speaker, independent of context. As a result, a great deal is known about how individuals store, organize, and access knowledge in the mental lexicon; how individuals parse sentences and resolve syntactic ambiguity; and how individuals plan and articulate utterances. But there is more to language processing than these (seemingly) autonomous processes, as has been demonstrated by those who work within the "language-as-action" tradition (e.g., Brennan & Clark, 1996; Clark, 1992; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1989, 1991, 1992; Glucksberg, Krauss, & Weisberg, 1966; Hanna, Tanenhaus, & Trueswell, 2003; Krauss, 1987; Schober & Clark, 1989). Consider three students, Leah, Dale, and Adam, who are trying to recall a scene from an excerpt of a movie[1] that they recently watched together, in which the protagonist is forced to wear an odd and embarrassing object:

> . . .
>
> Leah: um. . . then he gets punished or whatever?
>
> Dale: what was that, a wreath or—
>
> Leah: yeah it was some kind of browny—
>
> Adam: yeah it was some kind of straw thing or something
>
> Leah: mhm
>
> Dale: around his neck
>
> Leah: so that everybody knew what he did or something?

[1] The scene comes from a John Sayles movie, *The Secret of Roan Inish*.

Adam: straw wreath

Dale: yeah

. . . (excerpted from Brennan & Ohaeri, 1999)

Even though this transcript bears little resemblance to the idealized sentences typical of playwrights' scripts, psycholinguists' stimuli, or linguists' grammaticality judgments, it unfolds in an orderly way. The three partners rapidly succeed in establishing consensus as they share a focus of attention, cue one another's memories, and ratify one another's proposals about what to include in the product they are constructing together: their joint memory of the event. In doing this, they even complete one another's utterances. The product represented by this transcript reflects a process by which both memory recall and speaking are grounded in action conducted jointly, rather than achieved by minds working alone. Such data from studies of language-as-action (Clark, 1992; Tanenhaus & Trueswell, 2004) focus on language use in physical or communicative contexts. This particular spontaneous exchange comes from a large corpus recorded in an experimental study of collaborative recollection (Ekeocha & Brennan, 2008). It is so typical of everyday conversation as to seem rather unremarkable and yet at the same time, displays a level of coordination between partners that is astonishing in its virtuosity.

There is a growing trend within cognitive science to examine human cognition in social contexts, either pairwise or in small groups. This includes recall of memories (e.g., Ekeocha & Brennan, 2008; Harris, Paterson, & Kemp, 2008; Hollingshead, 1998; Weldon & Bellinger, 1997), collaborative visual search (e.g., Brennan, Chen, Dickinson, Neider, & Zelinsky, 2007; Neider, Chen, Dickinson, Brennan, & Zelinsky, 2005), decision making (e.g., Kiesler & Sproull, 1992; Wiley & Jensen, 2006), learning (e.g., Wiley & Bailey, 2006), two-person motor activities (e.g., Sebanz, Bekkering, & Knoblich, 2006; Sebanz & Knoblich, 2009), and of course, psycholinguistic processing in dialog. Some have argued that processing may be qualitatively different in the context of dialog than in monologue because both speech comprehension and speech planning systems are active at once (e.g., Pickering & Garrod, 2004). Others argue that, at least initially, language processes in dialog are identical to language processes in monologue because conversational partners process language from their own "egocentric" perspectives in which early processing is encapsulated from partner-specific information (e.g., Barr & Keysar, 2002; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998; Kronmüller & Barr, 2007), followed by a second stage in which they can take their partner's perspective into account. We take the view that processing in dialog can be explained by ordinary memory processes (Horton & Gerrig, 2002, 2005a, 2005b; Metzing & Brennan, 2003) and argue that these processes need not be encapsulated, but under some circumstances, are adapted flexibly and rapidly to the perspective of a conversational partner.

In addition to the coordination that takes place interpersonally, *between* partners, language processes are also coordinated intrapersonally, *within* the mind of an individual with many processes conducted in parallel: For instance, an individual speaker simultaneously plans and articulates an utterance while monitoring an addressee's reactions, and an individual addressee simultaneously listens to and interprets an utterance moment by moment while preparing what to say next, or even how to contribute to what the speaker is saying. This appears to require that various subprocesses of planning, parsing, interpretation, articulation, and monitoring must be able to share information and influence one another in a rather fine-grained way. Even though key capabilities that make human communication possible—such as the language faculty itself, the ability to mentalize about another person's mental state (or *Theory of Mind*—ToM), and the ability to respond rapidly and automatically to sensorimotor cues from human motion, speech, and other behaviors—may to some extent be supported by neural circuits thought to be distinct (Van Overwalle & Baetens, 2009), behavioral evidence suggests that there is close integration of these underlying processes (and their products), both within and between the minds of interlocutors. This, we argue, is what the study of language processing should aim to map, model, and explain.

In this chapter, we consider language processing in communicative contexts as a joint activity in which two or more interlocutors share or synchronize aspects of their private mental states and act together in the world. We summarize key experimental evidence from our own and others' research on how speakers and addressees take one another into account during communication. Under some circumstances, interlocutors can adjust to information about a partner's characteristics, needs, or knowledge in the early moments of processing. The accumulating evidence suggests that cognitive processing is probabilistic and flexible in how it adapts to partner-specific information (Brennan & Hanna, 2009; Jurafsky, 1996; MacDonald, 1994; Tanenhaus & Trueswell, 1995). We then discuss the evidence from electrophysiology and imaging studies that has begun to illuminate the neural architecture supporting joint and individual processing during communication.

2. Dialog: Beyond Transcripts

As evident from the example of the three students recalling a movie together, the process of coordinating meaning leaves behind striking evidence in the dialog transcript. A transcript is an analyzable product that can provide evidence about how interpersonal coordination unfolds, as one utterance seems to shape what is said next. Transcripts show that successive utterances produced by interlocutors often display recognizable contingency. One speaker may complete another's utterance by adding an installment that seamlessly continues its syntactic structure, as in our opening example (for studies of collaborative completions, see DuBois, 1974; Lerner, 1996; Wilkes-Gibbs, 1986). Many important descriptive insights about structural phenomena in conversation such as turn-taking, repair, and co-construction of utterances have been presented by ethnomethodologists who analyze detailed transcripts of naturally occurring conversations (e.g., Goodwin, 1981; Jefferson, 1973; Sacks, Schegloff, & Jefferson, 1974).

Although a transcript can be informative, it is only an artifact of the processes that generate it; people who overhear a conversation (including those who analyze it later) may not understand it in the same way that participants do (Kraut, Lewis, & Swezey, 1982; Schober & Clark, 1989). Psycholinguists who study dialog are interested in systematically probing the processes from which a transcript emerges. To understand what people might intend when they say what they say, psychologists (e.g., Clark, 1992; Glucksberg et al., 1966) have wrestled conversation into the laboratory in order to test hypotheses about language use and processing (often inspired by insights from conversation analysts). Experimental control and reliability are achieved by assigning different pairs of subjects to complete the same task in which they refer to, look at, pick up, and move objects. By observing such task-oriented dialog, the experimenter has access not only to the transcript, but also to physical evidence of what speakers mean and what addressees understand. This has led to conclusions about the underlying cognitive mechanisms of phenomena such as lexical choice and variability, perspective taking, distribution of initiative, conversational repair, the accumulation of common ground between partners, and audience design, or tailoring an utterance to a particular partner.

Consider these three excerpts from the transcript of a referential communication experiment in which two naïve partners could hear but not see each other (Stellmann & Brennan, 1993). Partners A and B each had a duplicate set of 12 cards displaying abstract geometric objects. The matcher (B) needed to arrange his cards in the same order as the director's (A's) cards. They did this for the first time in Trial 1, after which the cards were scrambled and matched again repeatedly (Trials 2 and 3):

Trial 1:

A: ah boy this one ah boy alright 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: alright 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: alright I got it (*etc. – they match about a dozen other cards*)

Trial 2:

B: 9 is that monk praying

A: yup (*etc. – they match other cards*)

Trial 3:

A: number 4 is the monk

B: ok

This matching task elicits data about interlocutors' spontaneous productions (from the transcript) and interpretations (from observing physical evidence provided by when and where the matcher moves the cards). The combination of behavioral evidence in the context of an experimentally controlled setting, synchronized with speech documented in the transcript, has provided powerful evidence for *common ground* or partially and mutually shared mental representations that presumably accumulate in the minds of both partners as they interact (whether in a laboratory experiment or in everyday conversation). *Grounding* enables partners to achieve a joint perspective on an object, such that referring to it becomes more efficient over time. The process of grounding typically results in *entrainment*, or convergence and synchronization between partners on various linguistic and paralinguistic levels—including in wording, syntax, speaking rate, gestures, eye-gaze fixations, body position, postural sway, and sometimes pronunciation (e.g., Branigan, Pickering, & Cleland, 2000; Brennan & Clark, 1996; Giles & Powesland, 1975; Levelt & Kelter, 1982; Shockley, Richardson, & Dale, 2009). Transcripts of different pairs of partners referring repeatedly to the same object demonstrate that there is less variability in the wording and perspectives associated with objects *within* a particular dialog than *between* dialogs (Brennan & Clark, 1996). In one experiment, 13 pairs each created, entrained on, and consistently reused one of 13 different perspectives for the geometric tangram figure in Figure 1 (Stellmann & Brennan, 1993).

The perspective that two interlocutors ground during a dialog, then, is another kind of joint product that emerges from interpersonal interaction. At the same time, interlocutors who share a communicative goal can be flexible in revising jointly achieved perspectives when necessary. And they can be extremely flexible in what they are willing to negotiate an expression or even a single word to mean.

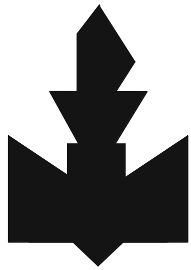

Figure 1 Perspectives vary across conversations.

Although a transcript can vividly illustrate some of these interpersonal products of interactive dialog, it often says little about how language processing unfolds incrementally and intrapersonally (within the mind of a participant). A major methodological advance has been the "visual worlds" paradigm pioneered by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995). This experimental paradigm measures the looking behavior of listeners who wear inobstrusive, head-mounted eye trackers while hearing prerecorded or scripted utterances that refer to visible objects; it measures indirect evidence of processing at a fine temporal grain, computed from the proportions of looks to an object within a defined epoch, in order to uncover the time course of lexical, prosodic, syntactic, semantic, and pragmatic processing (e.g., Altmann & Kamide, 2007). Some recent studies have merged the visual worlds eye-tracking paradigm with referential communication tasks done jointly by two spontaneously interacting partners (e.g., Brown-Schmidt, 2009; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Hanna & Brennan, 2007; Kraljic & Brennan, 2005). This approach has the potential to uncover not only how processing unfolds online within an individual engaged in dialog, but also how processing is coordinated incrementally between individuals.

3. Process Models of Dialog

What is the nature of dialog? All experimental studies of collaborative cognition rely on some notion, often entirely implicit, of what it means to participate in a dialog or to otherwise process information along with a partner (Kuhlen & Brennan, 2008). Some studies rely on the *mere presence* of one or more partners who may not be allowed to interact; this approach presumes that the effect of interpersonal collaboration is strictly motivational. Others allow a partner to contribute to the interaction only once, which decouples coordination processes from language processing. These approaches seem to assume that collaboration is based on a unidirectional exchange of information: While one conversational partner speaks the other listens passively. Some studies control the timing, order, or kinds of contributions that partners may make during a task (e.g., Basden, Basden, & Henry, 2000; Wright & Klumpp, 2004); while this may be desirable for controlling variation due to behavioral contingencies, it removes partners' ability to take initiative, treats what may be meaningful coordinating signals as noise, and probably rules out any but the simplest sorts of coordination of the processes under study.

Some psycholinguistic studies of dialog gain control by using confederates (whether human or simulated). But unless a confederate is doing the task for real, with actual communicative needs, the confederate's behavior can differ in troubling ways from the spontaneous behavior of a naïve partner. For instance, when a confederate plays the role of an addressee over and over in a study about speech production, she may know what the speaker is about to say better than the speaker himself does, and her feedback and nonverbal cues, if not carefully characterized and controlled, are very likely to communicate her *lack* of a need for information (Brennan & Williams, 1995; Kuhlen & Brennan, 2010; Lockridge & Brennan, 2002). For that reason, we are wary of using confederates in the addressee role unless they are actually doing a task with the subject. Most of our studies of language use and processing have used pairs of truly naïve speakers and addressees (e.g., Bortfeld & Brennan, 1997; Brennan, 1990, 1995, 2004; Brennan & Clark, 1996; Brennan & Ohaeri, 1999; Brennan et al., 2007; Ekeocha & Brennan, 2008; Galati & Brennan, 2010a; Hanna & Brennan, 2007; Kraljic & Brennan, 2005; Lockridge & Brennan, 2002). Some have had copresent confederate speakers who interact mostly spontaneously with naïve addressees, producing only certain critical utterances according to a partial script (e.g., Hwang, Brennan, & Huffman, 2007; Metzing & Brennan, 2003). A few have used prerecorded utterances but without any pretense that a live speaker is present (e.g., Perryman & Brennan, 2009). The point is that one partner's behavior shapes another's during dialog or during collaboration more generally (Kuhlen & Brennan, 2010), and this should be acknowledged when confederates are employed (Kuhlen & Brennan, 2008).

In this section, we describe three influential views of processing in dialog, each of which makes quite different assumptions about its essential aspects.

3.1. The Message Model

The message model of communication (or as Pickering & Garrod call it in their 2004 critique, the *autonomous transmission model*) is intuitively plausible and widely assumed among the cognitive sciences (e.g., Akmajian, Demers, & Harnish, 1987). This model is derived from information theory (MacKay, 1983; Shannon & Weaver, 1949; Wiener, 1965), in which information is defined in probabilistic terms; what is less probable is more informative. Communication involves the transmission and reception of information, which flows at a particular rate through a channel. One agent, a sender, encodes a message into a language and transmits it to another, a recipient, who decodes it; the two agents can communicate as long as they both have the same set of encoding and decoding rules (e.g., a language). Feedback (e.g., "backchannels" in conversation; Yngve, 1970) regulates the flow of information. The message model is consistent with the *conduit metaphor* (see critique by Reddy, 1979), in which words are treated like packages of meaning sent by speakers to listeners. It is difficult to think formally about communication without invoking the conduit metaphor and other information theoretic terms (Eden, 1983).

The approach represented by the message model decouples coordination from language *per se*, and it does not require that one partner recognizes an intention to communicate in the other. It has been used to model interactions between humans, between nonhumans, between mechanical processes, and between humans and machines (Wiener, 1965). But it is difficult to see how the message model could explain the tightly coordinated exchange among Leah, Dale, and Adam, in which their contributions defy relegating them to roles of sender or receiver, and meanings have no simple mapping but are negotiated so fluidly and flexibly. As these three recall the movie together, they coauthor a jointly recalled and articulated product (rather than formulating and sending signals autonomously). They all recognize a common goal. And in the first trial from the "monk praying" example, Partner A was the one who knew the identity of the target objects (and so should be considered to be the sender of the message), and yet it is B (the recipient) who ended up proposing the perspective that they entrain upon. As Figure 1 illustrates, there is no predictable mapping of perspective or label to object. We argue (as do Reddy, 1979; Schober, 1998) that words do not "contain" their meanings; even labels for common objects that are highly conventional can turn out to be negotiable. This means that there is no guaranteed 1:1 mapping of meaning to word, even for basic level terms. As Brennan and Clark (1996) showed in a series of referential communication studies, once speakers have entrained upon a perspective for a common object (e.g., calling a shoe *the man's loafer* to distinguish it from other shoes), they often continue to use the over-informative term even when this level of detail is no longer necessary (when *the man's loafer* is the only shoe). In fact, native speakers of English may even produce wildly nonidiomatic referring expressions (e.g., *the chair in which I shake my body* for a rocking chair or *the chair with five little tires on the bottom* for an office chair) to maintain a perspective that has been mutually achieved with a non-native speaker (Bortfeld & Brennan, 1997). The message model does not account for such flexibility. Because we are interested in understanding how people coordinate joint actions interpersonally and how they coordinate joint action with language processing intrapersonally, we find that the message model presents an unsatisfying view of communication.

3.2. Two-Stage Models

Several accounts of cognitive processing in dialog can be grouped together because they presume that processing is conducted in two distinct stages. According to the "interactive alignment" model (Pickering & Garrod, 2004), language processing in a dialog setting is fundamentally different from language processing in monologue because in dialog, both the speech production and speech comprehension systems are active at once, with the two systems assumed to have parity of representations. The interactive alignment model further assumes that interlocutors routinely come to achieve shared mental representations through a "direct" process of priming. Priming is proposed as the mechanism that explains convergent linguistic behaviors both between and within interlocutors such as lexical entrainment, shared perspectives, and the reuse of syntactic forms. According to this account, interlocutors converge on shared terms (such as in our earlier "monk praying" example) simply because one partner's utterance primes another's. *Interpersonally*, alignment is claimed to be direct and automatic. As the basis for such imitation, Pickering and Garrod (p. 188) invoke the human mirror system (to be discussed in Section 6), as well as the fact that the same brain areas (Brodmann's Areas 44 and 45; see Iacoboni et al., 1999) are implicated in both language processing and imitation. On Pickering and Garrod's view, processing in dialog defaults to what is assumed to be automatic and inflexible, driven by priming.

The interactive alignment model is compatible with two-stage proposals by Keysar and colleagues (e.g., the "monitoring and adjustment" theory: Horton & Keysar, 1996 and "perspective adjustment" theory: Keysar, Barr, & Horton, 1998) that assume that early processes in dialog are unable to take account of a partner. On these proposals, interlocutors often share the same context, knowledge, or informational needs, so that what *appears* to be audience design (when one partner seems to take the other's knowledge or mental state into account) is actually done for the self (Brown & Dell, 1987). As with the interactive alignment model, the first stage of these models is fast, automatic, and encapsulated from all but "egocentric" information, followed by an inferential stage that can accommodate partner-specific information, but more slowly. On these approaches, such mentalizing about a partner (or deploying "full common ground" to plan or process an utterance) is thought to be computationally expensive (e.g., Pickering & Garrod, 2004, p. 180), and therefore either optional or else invoked only when necessary for a repair: "normal conversation does not routinely require modeling the interlocutor's mind" (Pickering & Garrod, 2004, p. 180).

The interactive alignment theory further assumes that, *intrapersonally* or within the mind of an individual, priming at one level of linguistic processing (e.g., phonological) leads directly to alignment at another level (e.g., lexical representation), and that this automatically results in shared representations between partners at all levels of linguistic processing (Pickering & Garrod, 2004). But for this proposal to work, both interlocutors would have to be exact copies of one another. The problem is that presumably any conceptual networks that undergo priming within an individual's mind will have been sculpted by their idiosyncratic experiences and memories, and so it seems unlikely that shared meanings can be reached simply by priming (see Schober, 2004 for a related critique). Priming is simply the underlying currency by which language and memory are purchased, with multiple elements being primed at a given moment. As we will argue in Section 5.3, priming is *not* a satisfying explanation for convergent behaviors such as entrainment because such behaviors have a partner-specific component.

Note that not all of the theories that assign a prominent role to priming in order to account for convergent behavior agree that priming results in shared mental representations. In the "coordinative structures" proposal (Shockley et al., 2009), which focuses on convergent behaviors such as gaze patterns, body sway, and postural coordination, the authors argue that at least for these behavioral adjustments, executive control (and presumably mentalizing) does not play a role (p. 315) since these behaviors happen too rapidly, and since postural mimicry and sway are largely unconscious. The question remains, then, whether linguistic and communicative behaviors can also be aligned at multiple levels of linguistic processing without involving executive control and without achieving aligned mental representations.

3.3. The Collaborative View and the Grounding Model

Like the interactive alignment model, the grounding model views dialog as fundamentally different from monologue, but for different reasons (see Clark & Brennan, 1991 for discussion; see Cahn & Brennan, 1999; Clark & Schaefer, 1989 for formal models of grounding). According to this view, spoken communication is conducted not only as a kind of joint activity, but as a collaboration (Clark, 1992; Clark & Wilkes-Gibbs, 1986). On this view, words do not "contain" meanings, there are no "default" contexts, and entrainment and understanding are not automatic byproducts of priming. Rather, communicative signals are intended to be recognized as such by communicating partners. Meanings are coordinated through grounding, the interactive process by which people in dialog seek and provide evidence that they understand one another (Brennan, 1990, 2004).

Evidence used for grounding can be explicit, such as a backchannel response (*uhuh*) or clarification question, or it can be implicit, such as displaying continuing attentiveness via eye contact or continuing with a next relevant utterance. Interlocutors spontaneously provide evidence of what they themselves understand; they also monitor one another for such evidence, and when it is not forthcoming (or else not what they expect), they seek it out. Depending on their purposes and the task at hand, they set higher or lower grounding criteria for the form, strength, and amount of evidence they seek or provide at any particular point (Brennan, 1990, 2004; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs, 1986).

According to Clark and Schaefer's (1989) grounding model, Partner A cannot know whether her utterance (‘‘number 4 is the monk'') constitutes a contribution to the conversation (and to the common ground she is accruing with Partner B) until there is some evidence, verbal or nonverbal, about how (or whether) Partner B has heard and understood it (‘‘ok''). On this model, each contribution to a conversation has a presentation phase (an utterance) and an acceptance phase (the evidence that comes after it). A speaker evaluates her addressee's response against the response she expected; she can then refashion her utterance and represent it, or even revise her original intention so that it now converges with the one her addressee seems to be recognizing or proposing. Elsewhere we have conceptualized grounding as a process of *joint hypothesis testing* (Brennan, 1990, 2004), by which an addressee also forms incremental interpretations or meaning hypotheses as an utterance unfolds (Krauss, 1987) and then tests and revises them as more evidence accrues. From the speaker's perspective, the unfolding utterance embodies her hypothesis about what she believes might induce her addressee to recognize and take up her intention at a particular moment.

Experimental studies of grounding often observe pairs of interlocutors doing a joint task, such as matching duplicate objects (as with the three trials in our previous example in which Partners A and B became increasingly efficient while discussing tangram figures). What began as a provisional, complex, and possibly incoherent proposal for a suitable perspective on an object (Trial 1 in our previous example) was ratified during the grounding process; both partners converged on an efficient and streamlined label for a perspective built on their common ground (Trials 2 and 3). Both took responsibility for making sure communication succeeds, not just Partner A (the one who knew the target configuration):

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

B: yeah like like a monk praying or something

According to the assumptions of the message model, which assumes that communication is about one person who has information transmitting it to another who does not have it, this should not happen. According to the collaborative view, this is not unusual.

Sometimes it is not clear whether partner-adapted processing is due to cues produced during the grounding process, or from the explicit representation of a partner's perspective. An early study that documented partner-adapted referring during referential communication (Brennan & Clark, 1996) had pairs of naïve speakers establish referential precedents during spontaneous conversation (e.g., using *the high heel*, to distinguish one shoe from several); after that, speakers either continued to interact with the same partner or else were paired with a new one to match the same objects. When continuing with the same partners, speakers continued to use the same terms they had entrained upon even when this was over-informative (e.g., when there was only one shoe in the set). But they tended to switch to the unadorned basic level term (e.g., *shoe*) when interacting with a brand new partner who had not matched the objects before. This partner-specific effect may have been shaped by speakers mentalizing about what their partners knew, by cues that partners presented about their knowledge or needs during the dialog, or by both of these factors in combination. These two sources of information may be independent, or they may interact.

4. The Role of Cues in Grounding

Experimental work within the grounding framework has focused on coordination by examining the role of nonlinguistic and nonverbal cues, including elements that other traditions have considered mere noise—either a product not worth studying or one too difficult to study systematically. These elements include paralinguistic cues (both verbal and nonverbal) such as acknowledgments or eye contact (Schober & Clark, 1989). Paralinguistic cues may be used in a variety of ways, such as to display an addressee's continued attention to (or confusion about, or alignment with) an utterance, to signal a speaker's degree of commitment toward what she is saying, to invite an addressee to participate in completing an utterance, to capture the addressee's attention, to display a speaker's awareness of a speech disfluency or other problem in speaking, or to initiate or invite a repair (e.g., Brennan & Williams, 1995; Clark & Fox Tree, 2002; Goodwin, 1981). Additional evidence of a partner's understanding comes from incremental progress in whatever joint task interlocutors are doing (Brennan, 1990). During the process of grounding, interlocutors produce and monitor paralinguistic cues and monitor one another's instrumental behavior in order to seek and provide evidence that they understand one another.

We propose that the use of such cues in grounding facilitates the kind of intrapersonal "mind reading" needed for interlocutors to conclude that they are both talking about the same thing. These paralinguistic signals (*track 2* or secondary signals; Clark, 1994, 1996) provide information about the ongoing utterance itself (as distinct from *track 1* signals, which encode the "official business" of the utterance; Clark, 1994, 1996). The interactive alignment model (Pickering & Garrod, 2004), along with its cousins (Barr & Keysar, 2002; Dell & Brown, 1991; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998), ignores any early or automatic role that such cues may play in shaping language processing in dialog (largely ruling out the kind of flexible collaboration that such signals could help achieve, and instead focusing on what is achieved by automatic, "dumb" priming). Most versions of the message model allow a role for backchannel cues limited to regulating the rate of information flow rather than modeling how the evidence provided by a partner may collaboratively shape the incremental products of dialog. Of the models we have reviewed here, only the grounding model assigns a major role to such cues.

Are such cues really *communicative*? An essential aspect of communication is the ability of one person to recognize another's intention to communicate. This, according to Grice (1957), is what differentiates natural information (e.g., smoke is a *symptom* caused by fire) from non-natural (e.g., a smoke *signal* may be recognizable as an intentional communicative act). What starts out developmentally as a natural cue, such as a cry of pure distress produced by a baby who is hungry, develops into an intentional display intended to be communicative, as when a child cries to get her parents' attention. Although savvy parents can tell the difference, sometimes the distinction between natural and non-natural cues is ambiguous (see Harding, 1982 for more on relevant cues in development). A cue may serve both communicative and instrumental purposes; it is not always easy to differentiate communicative from noncommunicative behavior. Consider the production of *um* and *uh*, short elements sometimes known as "fillers." Clark and Fox Tree (2002) have argued that such signals are communicative, that they can facilitate processing, and in fact, that *um* contrasts with *uh* in much the same way that lexical items do. However, facilitation may be due to the time that elapses while the filler is produced rather than to its phonetic form (Brennan & Schober, 2001). Moreover, a cue can facilitate processing for an addressee without being communicative. Consider three criteria that must be met for a cue to be "communicative" (proposed by Brennan & Williams, 1995):

Criterion 1. The cue must be potentially informative; that is, it must encode information.

Criterion 2. The addressee must be able to process the cue and recover the information.

Criterion 3. Finally, the cue must be able to be modified by the speaker's intentions. This does not require that the speaker be consciously aware of planning or modifying the cue *per se*, but only that the cue be shaped by the speaker's intentions toward the addressee or what they are doing together.

We acknowledge that some paralinguistic cues may be produced communicatively while others may not be; nevertheless, even the cues that do not meet Criterion 3 can still serve a coordinating function, helping partners in conversation seek and provide evidence about what each other intends and understands.

Consider the phenomenon of "Feeling of Knowing" (Hart, 1965), the metalinguistic ability to assess one's own knowledge. Speakers can display their confidence (or lack thereof) when they answer a question, via the latency to their answer, the use of rising intonation, a filler such as *uh* or *um*, and self-speech (Smith & Clark, 1993). Speakers who display uncertainty while recalling an answer or certainty when saying "I don't know" are likely to fail to recognize the answer later on a multiple choice test. This satisfies Criterion 1; the paralinguistic cue displays reliable information about what the speaker really knows. It turns out that these cues are also interpretable by addressees (as a "Feeling of Another's Knowing," Brennan & Williams, 1995; Swerts & Krahmer, 2005), satisfying Criterion 2 and potentially aiding coordination. However, such cues may simply emerge from the speakers' own ease or difficulty in recalling, planning, and articulating an answer; whether they are actually communicative or not depends on whether speakers modify the cues based on their intentions toward their addressees. One way to test for Criterion 3 is to have speakers answer questions that are either sincere (the speaker knows that the partner who asked the question does not know the answer) or rhetorical (the speaker knows that the partner knows the answer, similar to a student answering a question posed by a teacher; Brennan & Kipp, 1996; Brennan, Kuhlen, & Ratra, 2010).

So far we have focused our discussion of cues on their potential as interpersonal signals in the process of grounding, as revealed in dialog transcripts. In the next section, we consider evidence for partner-specific impacts as revealed by the time course of eye gaze and other behaviors synchronized with linguistic evidence.

5. Partner-Specific Processing

It is clear from the evidence in a dialog's transcript that speakers tailor their utterances to what they know about addressees, and that addressees tailor their interpretations to what they know about speakers. What is not so clear is how and when they do this. The models of interactive communication described in Section 3 make very different predictions about partner-adapted processing.

Recall that according to the message model, processing language in dialog is not so different from processing in monologue; interlocutors take discrete turns, with one listening while the other is speaking and *vice versa*. Partner-adapted processing is not an issue because words map simply onto meanings; rules of encoding and decoding guarantee successful communication, as long as the transmission channel is not noisy or otherwise defective. The recognition of communicative intention is beside the point. According to the interactive alignment model, processing in dialog is distinctly different from processing in monologue, with an individual's production and comprehension systems both active at the same time during dialog, so that processing is assisted by an assumed parity between representations for speaking and representations for interpretation. One interlocutor's behavior primes another's, such that convergence of their mental representations is largely automatic. Like the two-stage interactive alignment model, the monitoring and adjustment model predicts that processing, at least initially, is automatic and inflexible; people with different perspectives or knowledge default to processing in a way that is not adapted to a partner, and they take account of "full common ground" only later (if ever), as a kind of slow inference or repair.

Grounding, on the other hand, assigns an essential role to recognizing and signaling communicative intent; dialog can be viewed as a highly coordinated hypothesis-testing activity that individuals engage in together, where one partner's presentation (their hypothesis of what their partner will understand) plays a dual role by providing the other person with evidence of how the previous utterance has been understood. Products such as utterances and perspectives are jointly constructed. This sort of model supposes that partner-specific processing is flexible and "smart," as well as highly incremental.

In Section 5, we consider experimental evidence about the products and timing of partner-adapted processing in dialog. We discuss some of our own and others' behavioral and eye-tracking data that are relevant to the agenda of uncovering a cognitive architecture that could support such effects.

5.1. Global and Local Adaptations

It is useful to categorize partner-specific information into two sources: (1) information from a more or less global model of a partner or their characteristics, mentally represented from prior personal experience, from expectations, or else from a stereotype, and (2) feedback that becomes available locally online, from cues that emerge as the dialog unfolds. The first source of information involves some degree of mentalizing about the partner and their intentions. It is available in some form at the start of the dialog (whether in detailed or else quite rudimentary form), and it may or may not be updated as the dialog unfolds. The second source consists of evidence emerging during the interaction about the context or the partner's needs, perceived from verbal and nonverbal cues. Whether a particular kind of cue evokes mentalizing, and when such mentalizing might occur, depends on the attributions made to the cue (as we will see presently). Presumably if a cue satisfies all the criteria to be considered as communicative (including being able to be mediated by intention, as outlined in Section 4), mentalizing is involved; if the cue satisfies only the first two (is informative and can be perceived), then it may support interpersonal coordination but not involve mentalizing.

Both global and local sources of partner-specific information have the potential to guide production of utterances. In one study (Brennan, 1991), students were led to believe they were interacting via text with either a remotely located student or else a computer that could interpret natural language; the task was to retrieve information to fill in the missing cells of a spreadsheet database about hypothetical students and their characteristics. The answers were provided by a confederate (blind to whether she was assumed to be human or computer), were entirely rule-based, and in a given dialog, took the form of either short elliptical and telegraphic turns, or else complete sentences that reused syntax and word choice from the students' original questions. Those who believed they were communicating with a natural language interface began the dialogs by typing telegraphic utterances, whereas those who believed they were communicating with a remotely located person began with longer, grammatical sentences. But this global force for audience design was trumped midway through the session by the remote partner's online feedback; by the end of the sessions, students' questions converged in form with their partners' answers (to either short utterances or complete sentences), regardless of whether the partner was believed to be human or computer. Although this pattern of adaptation was true for some kinds of measures (e.g., lexical choice and syntactic form), it was not true for all measures. For instance, students used third-person pronouns relevant to the task equally often in all conditions (e.g., *Where does **he** work?*), showing that they expected their (human or computer) partner to model connectedness of utterances within the dialog context, but they rarely used first- or second-person pronouns with computer partners compared to with humans (e.g., *Can **you** tell **me** whether...?*), suggesting that they did not expect to have social context with computers.

Often, local cues (e.g., feedback about the informational needs of a conversational partner) corroborate the information available through global cues (e.g., about a partner's identity). This can make it challenging to tease apart effects of these two potentially independent factors, and most studies do not attempt to do so. In a recent study (Kuhlen & Brennan, 2010), we teased apart expectations about a partner from cues. Speakers learned jokes in the form of brief stories and told them to addressees who also were naïve subjects. The instructions led speakers to expect either attentive addressees (who would have to retell the jokes later), or distracted addressees (working on a secondary task while listening to the jokes). As expected, attentive addressees gave more feedback than distracted addressees. Thus, while (globally) expecting attentive or distracted addressees some speakers encountered behavior contrary to their expectation (based on local cues in form of addressee feedback). We found that the tellings of the jokes were shaped both by speakers' expectations and by addressees' cues. Speakers with attentive addressees told the jokes with more vivid detail than those with distracted addressees, but only when they *expected* attentive addressees. Speakers with distracted addressees put less time into the task than did those with attentive addressees, but only when they had expected the distracted addressees to be attentive (when the initial expectation did not match the unfolding evidence). These results suggest that feedback cues are interpreted against prior expectations or attributions about a partner.

A similar pattern of partner-specific adaptations was found in speakers' speech-accompanying gestures (Kuhlen, Galati, & Brennan, 2010). Independent of adjustments made in speaking, speakers gestured more frequently when their expectations were consistent with addressees' feedback, supporting the idea that speakers put more effort into narrating when their global expectations of addressees' needs are matched by local cues provided by addressees in the interaction. Moreover, speakers used more gestures that were produced in the body's periphery when narrating to attentive addressees whom they had also expected to be attentive, supporting the idea that consistency between local and global cues is associated with more vivid narration. These results suggest that global information established prior to the interaction is updated by local cues provided within the interaction in a highly interactive manner, resulting in a cascade of adjustments in speakers' narrating style that affects both speech and gesture.

A clear example of cues intended by one partner to be recognized by the other as communicative (and recognized by the other partner as such) comes from Brennan's (1990) study (reported in Brennan, 2004). Pairs of subjects in adjoining cubicles discussed target locations on identical maps displayed on networked computer screens. The task was for the matcher to get his car icon parked in the same target location displayed on the director's screen. In one condition, the director could visually monitor the progress of the matcher's car; in the other, she could not. In both conditions, they could talk freely; in both, the matcher saw only his car icon displayed over the map. Over 80 trials with different targets, whether the director could see the matcher's movements toggled every 10 trials (and the matcher was informed of this switch at the start of each block of 10 trials). So the director had local cues of what the matcher understood, updated moment by moment, while the matcher had only global information (that he needed to keep in memory) about what his partner could see.

When they could not visually monitor the matchers' progress in the task, directors proposed descriptions in installments, and matchers responded verbally to clarify, modify, and eventually, ratify descriptions of the target location. Meaning was established incrementally and opportunistically, with both partners sharing the responsibility for doing so (as with the earlier dialog about the tangram that looked like a monk). The matcher's icon typically arrived at the correct target location early in the trial; but they still needed additional verbal turns during which they grounded their meaning. It was up to the matcher to propose when he thought he understood well enough for current purposes and go on to the next trial. In contrast, when the director could monitor the matcher's icon's movements, she took the responsibility for determining when the matcher indeed understood the target location, and since this was based on direct visual evidence, she took responsibility of proposing when to go on to the next trial, sometimes suspending speaking midword as soon as the matcher reached the target, as here (note: asterisks denote overlapping speech):

Director: ok
    now we're gonna go over to
    M-Memorial Church?
    and park right in Memor-
    right there
    that's *good.*
Matcher: *that's* rude
    to park in the church.
Director: hheh heh

Grounding with visual evidence was much more efficient, although partners adjusted their effort so that performance was equally accurate with and without visual evidence. What is particularly striking is that even though matchers' screens appeared the same to them regardless of what condition they were in (there were no cues to remind them of what directors could see), they easily adapted to what they knew about their unseen partners' perceptual context by providing or withholding back-channels; when they knew the directors could see their cars, they used their icon moves not only as instrumental acts for doing the task, but also as communicative acts (Brennan, 2004). Each time the visual evidence condition toggled, matchers adapted to this global partner-specific information immediately (almost always without discussion). Directors packaged location descriptions into installments and grounded these with the online local cues provided by matchers' icon movements. So in this study, directors used local cues provided moment by moment by their partners; these were verbal when they could not see their partners' moves, and visual when they could. At the same time, matchers, who were aware when their moves could or could not be seen, used that simple bit of information to guide whether to produce backchannels or not.

5.2. Speakers Adapt Utterances for Their Addressees

Interlocutors often share considerable context beyond being speakers of the same language, including that due to previously established common ground or to being copresent in the same perceptual environment. Therefore, what might *appear* to be a case of a speaker tailoring an utterance to an addressee's needs or knowledge may occur simply because that is what is easiest for the speaker to do. For example, within a discourse the first articulation of a word (when it represents new and sometimes unpredictable information) tends to be longer in duration and more intelligible than repeated mentions of the same word (or other uses in which it is more predictable) (Bard et al., 2000; Fowler & Housum, 1987; Lieberman, 1963; McAllister, Potts, Mason, & Marchant, 1994; Samuel & Troicki, 1998). Listeners can pick up on attenuation as a marker of information status, such that when they hear an initially ambiguous word that is destressed, they assume that it refers to the *given* item in an array (that also includes a *new* distractor with the same phonological onset); but when the word is stressed, they assume that it refers to the new item (Dahan, Tanenhaus, & Chambers, 2002). The question is whether variations due to attenuation are communicative, for the benefit of the addressee (as assumed by Nooteboom, 1991; Samuel & Troicki, 1998), or whether this is a generic sort of variation produced automatically by speakers (Dell & Brown, 1991) that would likely occur without any addressee present. To establish that a variation in speaking is not egocentric but is produced truly as a form of audience design, "for" a partner, the perspectives of the speaker and the addressee must be distinguishable (for discussion, see Keysar, 1997; Lockridge & Brennan, 2002). Moreover, the speaker must be aware of her addressee's distinct perspective or needs in time to incorporate this information into speaking; if relevant information about the addressee's distinct perspective is not available in time, then a failure to incorporate it does not constitute a fair test of whether the early stages of speaking are egocentric (Horton & Gerrig, 2005a, 2005b; Kraljic & Brennan, 2005).

When telling stories, speakers leave out some details and include others; for example, they are more likely to mention atypical instruments and omit typical ones (which are implicitly associated with a particular verb or situation). A study by Brown and Dell (1987) tested whether this typicality effect is egocentric, or else driven by the needs of particular addressees. Eighty speakers read silently and then recounted aloud to a confederate addressee very short stories in which an instrument (either typical or atypical in association with a main verb) played a key role; the confederate either had or did not have a picture illustrating the main action and instrument (and the speaker subject knew what the addressee could see). Whether the addressee could see the instrument or not had no effect on whether and how speakers mentioned it; Brown and Dell concluded that the typicality effect was not an adjustment to the addressee's needs, but simply automatic for the speakers. However, their addressees (both of them) heard the same stories over and over, so actually knew them better than the speakers did; it is possible that the cues they provided signaled this. A subsequent study by Lockridge and Brennan (2002) had speakers tell similar stories, but to *naïve* addressees who had never heard the stories before, and who saw or did not see the pictures. Speakers were more likely to mention atypical instruments, to mention them early (within the same clause as the action verb), and to mark them as indefinite, when speaking to addressees without pictures than to addressees with pictures. This suggests that when addressees have real needs (and presumably signal them somehow), speakers take this into account in the syntactic choices they make early in an utterance (Lockridge & Brennan).

In another study, we examined the extent to which speakers attenuated elements of a longer story "for" themselves or "for" their addressees (Galati & Brennan, 2010a). Twenty naïve speakers spontaneously told and retold the same Road Runner cartoon story twice to one naïve addressee and once to another (counterbalanced for order: Addressee1/Addressee1/Addressee2 or Addressee1/Addressee2/Addressee1). This design enabled us to tease apart tellings of the story that were new versus old to speakers from those that were new versus old to addressees. We found that attenuation was mainly due to whether the material was new or old to the addressee rather than to the speaker; stories retold to the same (old) addressee were attenuated compared to those retold to the new addressee. This was true for a variety of linguistic units, including number of words, amount of detail, and number of events realized in the stories. Although lexically identical expressions by a same speaker were no different in length when addressed to a new versus an old addressee, expressions that had been addressed to new partners were more intelligible to a later group of listeners than when they had been addressed to addressees who had heard them before. This study provides strong evidence that attenuation is driven at least in part by the needs of addressees (in fact, it found little if any evidence for speaker-driven attenuation). The findings contrast sharply with that of Bard et al. (2000), who found that attenuation in articulation of repeated expressions depended on speakers' experience rather than addressees' (although it should be noted that their study did not tease apart speakers' from addressees' perspectives; all addressees where hearing the expressions for the first time).

We found a similar pattern of partner-specific attenuation in these speakers' gestures (Galati & Brennan, 2010b). Speakers produced fewer representational gestures overall in retellings to old addressees than to new addressees. The gestures produced in stories retold to old addressees were also smaller and less precise than those retold to new addressees (a for-the-addressee effect), although gestures were also attenuated over time (the only comparison from this experimental corpus that showed any for-the-speaker effect). These data support the conclusion that gesture production is guided by *both* the needs of addressees and automatic processes by which speakers do what is easiest for themselves.

Although Bard et al. (2000) (in their measures of duration and intelligibility) found no audience design effect at the grain of pronunciation of repeated words, Bard and Aylett (2000) did find audience design at the grain of referring expressions; their speakers marked expressions as definite when appropriate given the addressee's knowledge. The authors proposed a "dual-process model" in which automatic processes are modular and cannot take partner-specific context into account while other, more flexible processes can. But given the audience design effects on articulation that we found in Galati and Brennan (2010a), the modularity claim seems hard to defend. It may be that audience design effects on articulation are either produced inconsistently or that they are difficult to detect. On the other hand, a pattern of variable findings would be consistent with a system whose architecture allowed information to be incorporated into planning in a probabilistic (constraint-based) fashion (e.g., Jurafsky, 1996; MacDonald, 1994; Tanenhaus & Trueswell, 1995). A claim of modularity based on a null finding of audience design might be convincing if every stone has been overturned, and if the information in question is available early enough to impact planning (for discussion, see Brennan & Hanna, 2009; Kraljic & Brennan, 2005).

Variability in pronunciation is influenced by multiple factors. Hwang et al. (2007) examined the extent to which articulation may be governed by priming as well as by a conversational partner's communicative needs, using Korean-born speakers of English as a second language (L2). Ambiguities arise when non-native speakers fail to make L2 phonetic contrasts that are absent in their native language (L1). Korean speakers lack the voicing contrast b/p ("mob" vs. "mop") and the vowel contrast ae/E ("pat" vs. "pet"), so that when they speak Korean-accented English, the first words in each of these pairs are likely to be neutralized to sound like the second words. In two referential communication experiments, subjects who were Korean speakers of English spontaneously produced target words (e.g., "mob"). A confederate partner either primed the target words with a rhyming word (e.g., asking "What is below hob?") or did not prime them, and the referential contexts required pragmatically distinguishing two contrasting words ("mob" adjacent to "mop" in the array), or did not. The Korean speakers produced more English-like phonetic targets in *both* the priming and pragmatic conditions (vowel duration was used to signal both contrasts). Moreover, Korean speakers were primed to make the disambiguating contrast when interacting with an English speaker but not with another Korean speaker of English. These results show that Korean speakers speaking English (L2) can be led to produce a phonetic contrast that they do not have in L1 both when they are primed to do so *and* when their addressees need them to do to resolve an ambiguous expression.

Sections 5.1 and 5.2 have reviewed some studies of audience design in which interlocutors with distinct perspectives incorporate their partners' knowledge or needs rather than ignoring them or taking them into account at a late stage of processing. But we have not yet addressed the question of how perspectives (whether of self or other) are suppressed, selected, or updated moment by moment.

5.3. Addressees Adapt Utterance Interpretations to Speakers

According to the grounding framework, just as speakers design utterances for their addressees, addressees interpret utterances in the context of what they know about speakers. This means that the same words may be interpreted differently depending on who utters them. In a referential communication experiment that incorporated interaction between confederate speakers and naïve addressees, addressees' initial looks to familiar target objects (that they had previously grounded during interaction with a speaker) were delayed by a few hundred ms when the same speaker uttered an entirely new expression for the familiar object, but not when a new speaker uttered the same new expression (Metzing & Brennan, 2003). The conclusion was that speakers and addressees ground "conceptual pacts" or shared perspectives that are not only partner-specific but also quite flexible: Addressees were quick to abandon the precedent of a familiar expression when interacting with a new speaker; their first looks to the target were not delayed when the new speaker used the new expression.

This finding, that addressees experience interference or slowed processing when a conceptual pact (previously grounded with a particular speaker) is broken, has been replicated with young children, who show the effect when a speaker abandons a precedent for a new term without any apparent reason, but not when a new speaker introduces a new term (Matthews, Lieven, & Tomasello, 2008). These findings and related findings (e.g., Brown-Schmidt, 2009; Nadig & Sedivy, 2002) are incompatible with interactive alignment theory that seeks to explain convergence from priming alone (Pickering & Garrod, 2004), and in which the speaker's identity should not matter. Addressees do not inflexibly map expressions onto referents; within a pragmatic context (Grice, 1975), the identity of the speaker can be part of what is represented.

Finally, global information (specifically about a speaker) can interact with local information (from cues that emerge during dialog or speaking). That is, listeners interpret cues against the attributions that they make about those cues (Kuhlen & Brennan, 2010). For instance, when listeners hear a speaker's disfluency just before a referring expression, they interpret it online as evidence that the speaker is in the process of saying something difficult (Arnold, Tanenhaus, Altmann, & Fagnano, 2004)—*unless* they have a stable attribution for the disfluency (*the speaker has agnosia*; Arnold, Hudson-Kam, & Tanenhaus, 2007).

5.4. Simple or "One-Bit" Partner Models

It may be no coincidence that experiments that show audience design early in processing involve partner-specific information that is not only clear, but also already-computed and quite simple. In such experiments, what a partner needs is often captured by only two alternatives: *my partner can see what I'm doing*, or not (Brennan, 2004; Nadig & Sedivy, 2002); *my partner can reach the object she's talking about*, or not (Hanna & Tanenhaus, 2004); *my partner has a picture of what we're discussing*, or not (Lockridge & Brennan, 2002); *my partner and I have spoken about this before*, or not (Galati & Brennan, 2010a; Matthews et al., 2008; Metzing & Brennan, 2003); *my partner is currently gazing at this object*, or not (Hanna & Brennan, 2007); *my partner needs to distinguish this referent from a competitor*, or not (Hwang et al., 2007); *my partner is a young child*, as opposed to older (Shatz & Gelman, 1973); or *my partner is a native speaker of English*, or not (Bortfeld & Brennan, 1997). In these situations, an interlocutor may represent information in working memory about a partner's state as a simple either/or cue that can be flexibly updated as the situation changes. The findings of audience design in these situations demonstrates that a "partner model" need not entail a detailed record of all of the knowledge one partner has about what the other is likely to know (as well as what the other does *not* know, as pointed out in a critique by Polichak & Gerrig, 1998). In contrast, a simple "one-bit" model that does not require complex inferences or elaborate maintenance or updating could facilitate rapidly partner-adapted processing, even when two partners have distinct perspectives (Brennan & Hanna, 2009; Galati & Brennan, 2010a).

In the next section, we consider evidence from brain imaging studies about the neural circuits that may support partner-adapted processing, both by interpreting local cues and by maintaining simple models of interlocutors' intentions, perspectives, or communicative needs.

6. Neural Bases of Partner-Adapted Processing

Our cognitive/behavioral research program has followed the assumption that partner-specific adaptation during communication can be explained by general principles of memory and cognitive processing, rather than by special cognitive modules that either give priority to an egocentric perspective (Horton & Keysar, 1996; Keysar, Barr, Balin, et al., 1998; Keysar, Barr, & Horton, 1998; Keysar et al., 2000; Pickering & Garrod, 2004) or automatically restrict referential interpretation to what is in common ground (a position attributed to Clark & Carlson, 1981 by Barr & Keysar, 2002). Our studies and others that allow for spontaneous interaction between interlocutors (e.g., Brown-Schmidt, 2009; Brown-Schmidt et al., 2008; Hanna & Brennan, 2007; Kraljic & Brennan, 2005) demonstrate that partner-specific effects can emerge early in processing, and show no evidence for modular or two-stage (early egocentric, late partner-specific) processing models. We find that the evidence supports a cognitive architecture for language processing and communication that combines the available information in a parallel, constraint-based, and probabilistic fashion (Brennan & Hanna, 2009; Horton & Gerrig, 2002, 2005b; MacDonald, 1994; Metzing & Brennan, 2003; Tanenhaus & Trueswell, 1995).

However, the behavioral evidence does not tell us precisely how such flexible, partner-adapted processing is achieved in the brain. Imaging studies have revealed multiple neural circuits that appear to aid and abet everyday communication. These circuits handle a wide variety of cues and functions. Cues relevant to communication include gesture, eye gaze, nonlinguistic verbal cues, contrastive stress and other prosodic cues, and disfluencies. Relevant functions that may make use of these cues include speaking, linguistic parsing, postural and motor coordination during joint action, monitoring a partner's orientation or attention, evoking person stereotypes and other world knowledge, and last but certainly not least, mentalizing about their intentions or beliefs (Theory of Mind). Mapping the circuits that underlie these functions and discovering how these functions could work *together* requires deploying cognitive/behavioral tasks that preserve the essential aspects of communication. In this section, we discuss some recent and intriguing findings about the neural underpinnings of language and communicative processing that are potentially relevant to a more complete account of adaptive processing.

6.1. Mirroring

The idea that the production of speech relies on the same motor routines and representations as the interpretation of speech has been around for a long time (Galantucci, Fowler, & Turvey, 2006; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Over the past decade and a half, much evidence has accumulated that people perceive and understand the actions of others by relying on their own motor routines, using a common coding for both. Individual mirror neurons, activated both when an action is performed and when it is observed, have been identified in primates (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) and are presumed to exist in humans (Iacoboni et al., 1999). The human "mirror system" comprises a network that includes areas in the premotor cortex (PMC) and parietal cortex (in particular, the anterior intraparietal sulcus, aIPS), with input from the posterior superior temporal sulcus (pSTS) (Van Overwalle & Baetens, 2009). The existence of a more or less direct perception–action link is proposed to help people detect each other's goals by helping them *simulate* another's state, as "nature's way of getting the observer into the same 'mental shoes' as the target" (Gallese & Goldman, 1998, p. 497).

As Sebanz and Knoblich (2008) have pointed out, the mirror system has been misunderstood by some as being an inflexible mechanism that automatically supports mimicry, and hyped by others as being the explanation for all of social cognition. Recent accounts by these authors and others (e.g., Bekkering et al., 2009) argue that the truth is somewhere in between: the mirror system is recruited for rapid processing of a wide variety of cues and provides input to many kinds of processes, including those that support language, communication, and other forms of joint action. The perception–action links that the mirror system provides do not support only mirroring; arguably, most of the actions that people do jointly involve complementary or noncongruent actions rather than imitative or congruent ones, and the mirror system is more active during the preparation of complementary than imitative actions (Newman-Norlund, van Schie, van Zuijlen, & Bekkering, 2007). Currently, the value of the mirror system is presumed to be in facilitating the recognition of a partner's goal and the monitoring of outcomes of actions, rather than literally reproducing specific action primitives; otherwise it would be of limited use, as much of the time people would not be well served by an imitation reflex and since the mapping of action to goal is not 1:1 (for discussion, see Bekkering et al.; Van Overwalle & Baetens, 2009).

In this more flexible role—as a way to understand (but not mimic) the perspective of another by simulation—the mirror system could support partner-adapted processing by monitoring cues about a partner or about the objects in a task and rapidly updating the state of a simple or "one-bit" partner model. This would explain why interlocutors sometimes adapt rapidly to their partners' needs or knowledge rather than defaulting to behavior that appears to be egocentric.

6.2. Theory of Mind

Designing an utterance or action with regard to what a partner knows, or recognizing an utterance or other action as communicative (Grice, 1957, 1975), presumably involves mentalizing, or attributing intention to another.

Mentalizing involves neural circuitry that is usually thought to include (1) the medial prefrontal cortex (mPFC),[2] (2) the bilateral temporoparietal junction (TPJ),[3] and (3) the precuneus (BA 7)[4] (e.g., Ciaramidaro et al., 2007; Van Overwalle & Baetens, 2009; Vogeley et al., 2001). These areas are often considered to be core parts of a ToM network, activated during tasks that require taking into account another person's mental state. The classic ToM task (as tested on children in various stages of development) involves having a child witness an actor learning of the location of a hidden object, witness the object being rehidden in a different location unknown to the actor, and then predict where the actor will look for the object (Wimmer & Perner, 1983). The majority of imaging studies that aim to probe the ToM network in adults have been conducted in noninteractive settings in which subjects in an fMRI scanner read text stories about characters' true or false beliefs, perspectives, intentions, or motivations, compared to texts about characters' physical characteristics or objects (that do not require ToM to understand; Saxe & Kanwisher, 2003).

Generally speaking, mentalizing has been proposed to be a fast, automatic process rather than a slow, inferential one (Kampe, Frith, & Frith, 2003; Scholl & Leslie, 1999). This is consistent with the behavioral findings reported earlier, that interlocutors adapt processing to their partners from the early moments of processing. Within the ToM mentalizing circuit, the TPJ appears to be implicated whenever there are early and automatic inferences about another's goal, with the mPFC implicated during inferences about another's traits that unfold more slowly (Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009; Van Duynslaeger, Van Overwalle, & Verstraeten, 2007; see Van Overwalle & Baetens, 2009 for discussion). So ToM as a network may underlie not only immediate partner-adapted processing (in the TPJ region), but also the slower, inferential, adjustments to a partner that may unfold after an initially "egocentric" response (in the mPFC).

6.2.1. Distinguishing Kinds of Intentions: Private, Social, and Communicative

Some of the variability in findings about the shape of the network hypothesized to underlie ToM may be due to lack of precision in fMRI imaging, and some may be due to limitations in the kinds of intentions depicted in the stimulus stories. A study by Ciaramidaro et al. (2007) took a nuanced look at the neural bases of ToM, by having individuals in the scanner read short comic strips that distinguished (1) the private intentions of characters from their social intentions toward other characters, and (2) within these social

[2] Some studies label this Theory of Mind area as the anterior paracingulate cortex within the mPFC.

[3] Although note that some studies label this as the posterior STS, which extends to the TPJ.

[4] Some studies (Gallagher & Frith, 2003; Gallagher et al., 2002; Kampe et al., 2003) implicate the temporal poles (BA 38) in the ToM circuit.

intentions, communicative from noncommunicative intentions. The ToM areas mPFC and TPJ were both found to be crucial, but were activated differentially depending on the kind of intention being recognized. The right TPJ and precuneus were active in the processing of all types of prior intentions, with the anterior paracingulate cortex in the mPFC and the *left* TPJ active when processing social intention; in fact, in these comparisons the left TPJ was active *only* when processing communicative intention. The evidence from this study suggests four (rather than three) core parts for the ToM network, with distinct roles for both the left and right TPJ areas. It is possible that previous studies that failed to find a clear role for the left TPJ during mentalizing used stimuli that did not require recognizing communicative intentions; the authors suggest that the left TPJ may be a fourth ToM area activated by the recognition of intentions that are specifically intended as communicative.

6.2.2. Joint Activation During Interpersonal Interaction

While mentalizing about the intentions of characters in a story almost certainly overlaps with the mentalizing involved in thinking about an interlocutor's knowledge or communicative needs, a reading task probably misses some of the essential aspects of interacting with a partner in dialog. For instance, most ToM stimuli texts are written about characters in the third person rather than the first or second person, and most fMRI scanner tasks do not probe contingently unfolding social interaction between partners, with a few notable exceptions.

A few studies have used interactive games with real or simulated partners. In one series, neural activation was examined while pairs of partners playing a ''tacit communication game'' (Noordzij et al., 2009) in which ''senders'' invented new ways of conveying their communicative intentions to ''receivers'' using entirely graphical means. Senders had to figure out how to move icons so that receivers could distinguish instrumental moves from moves intended to instruct them about where they should move their own icons. The perspectives of the two partners were known (by both) to be different, with one person's icon being inherently more ambiguous than the other (a triangle that could be oriented in three ways, a rectangle that could be oriented in two ways, or a circle for which orientation did not matter). In this task, communication was interactive, incremental, and graphical; both communicative and control trials evoked identical motor actions and graphics so that activation related to the planning and interpretation of communicative intent could be distinguished from that related to noncommunicative signals, visual motion, and hand movements. In each session, fMRI data were collected from either the sender or the receiver. Remarkably, during communicative trials senders and receivers *both* showed activation in one of the same ToM regions: the right pSTS, but not in the left pSTS. This right activation was modulated by the degree of ambiguity in the communicative signal (e.g., a sender's circle could not easily depict how a receiver should orient their own triangle), but not by visual appearance or sensorimotor complexity. In addition to the right pSTS, mentalizing about communicative intent coactivated the mPFC. That the same ToM circuitry implicated in recognizing a partner's (the sender's) intention is also implicated in predicting how best to signal one's own intention to a partner (the receiver) suggests that there is a kind of functional parity between signaling one's own and interpreting a partner's intentionality.

One puzzle in comparing this study with the previous one (Ciaramidaro et al., 2007) is that Noordzij et al. (2009) reported no differential activation whatsoever for communicative action in the left pSTS (which extends into the left TPJ, the region where Ciaramidaro et al. *did* find activation associated with communicative intention). Whether this apparent inconsistency is due to a task difference remains to be settled. Noordzij et al.'s interactive task differentiated first- and second-person communicative intentions from instrumental acts, whereas Ciaramidaro et al.'s reading task differentiated third-person communicative intentions from other (ToM-associated) intentions. The interactive task required participants to generate communicative intentions as well as to recognize them, whereas they needed only to recognize them in the reading task. And the interactive task used graphical communication, whereas the reading task used language. There are so few imaging studies of communicative intention that it is difficult to interpret the implications of these task differences, but one speculative possibility is that the left TPJ might link ToM activation to language processing networks in the left temporal lobe.

6.2.3. Interactions with Human Versus Computer Partners

ToM is associated with predicting the behavior of conspecifics (e.g., Ciaramidaro et al., 2007; Van Overwalle & Baetens, 2009). But does it matter whether an interacting partner is human or computer? Several imaging studies have been conducted using tasks in which subjects interacted with computers or human partners (or else ones they believed to be human) in a prisoner's dilemma or other payoff game. In one such investigation (Gallagher, Jack, Roepstorff, & Frith, 2002), subjects who believed they were playing a (competitive) rock–paper–scissors game with either a computer or another person showed more activation in only one of the ToM areas with human than computer partners, the anterior paracingulate cortex (mPFC). In another investigation (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004), subjects playing interactive games and receiving feedback from supposed partners showed activation in two of the main ToM areas, the mPFC and posterior STS; these areas were activated whether subjects believed their partners were human or computer. The cues that subjects received during the sessions were identical (and automatically generated) in both partner conditions, and ToM was activated in both kinds of sessions, but activation was higher when subjects believed they interacted with humans (Rilling et al., 2004). This difference in activation may reflect partner-adapted processing that distinguishes human from machine partners, or it may emerge simply from different levels of engagement in the task; but either way, it documents the influence of the same sort of global partner-identity variable that has emerged in behavioral studies (e.g., Brennan, 1991). Recent studies by Krach et al. (2008, 2009) have found consistent results, with activation in the mPFC and right TPJ when interactive games were played with (supposed) human or computer partners; however in Krach et al. (2009), the first of these ToM areas was more activated when the partner was believed to be human than computer. When people played with one of four kinds of (simulated) partners (human, anthropomorphic robot, functional robot, or computer process), there was more activation in both of these ToM areas, the more human-like the partner (Krach et al., 2008). So the difference between interacting with a human partner and a computer partner may be quantitative rather than qualitative (at least for part of the ToM network). These studies suggest to us that under some circumstances ToM processing may be flexible enough to be able to model varieties of an intelligent partner's "mind" that need not even be human, an idea relevant to the field of "intelligent" computer–human interaction (Don, Brennan, Laurel, & Schneiderman, 1992).

6.3. Distinguishing a Partner's Perspective from One's Own: The Role of Executive Control

Stimulus stories that require recognizing a single character's intention presumably require less complex mentalizing than referential communication studies that require distinguishing two perspectives (e.g., one's own from one's partner's or one's private knowledge from common ground shared with a partner), especially when the two perspectives may in fact be inconsistent (Galati & Brennan, 2010a; Hanna et al., 2003; Metzing & Brennan, 2003; Nadig & Sedivy, 2002). Distinguishing privately held information from common ground presumably requires such mentalizing, as well as executive control to select the appropriate perspective and/or to suppress the inappropriate one. In addition, during dynamic communicative interaction, there is the challenge of keeping track of how a partner's perspective (or else common ground) changes over time.

Imaging studies show that the mentalizing network is recruited when people explicitly prevent themselves from imitating another's behavior (Van Overwalle & Baetens, 2009), perhaps facilitating the differentiation of self from other (Brass, Derrfuss, & von Cramon, 2005; Brass, Zysset, & von Cramon, 2001; for discussion, see Van Overwalle & Baetens). A study by Vogeley et al. (2001) attempted to distinguish egocentric (SELF) processing from ToM by comparing activation associated with stories about the intentions of another person to stories about the reader's own perspective. Consistent with other studies, Vogeley et al. found ToM to implicate the mPFC.[5] But reasoning about one's *own* perspective led to additional activation in the right inferior temporoparietal cortex that did *not* appear to be associated with ToM (Vogeley et al., 2001). These authors conclude that the right TPJ "is involved in computing an egocentric reference frame" (p. 179), and that ToM and SELF interact in the right prefrontal cortex, an area that has been associated with executive control processes.

To the extent that taking another's perspective requires inhibiting one's own, executive control seems to play a role by inhibiting responses that are either overlearned or imitative (Brass et al., 2005). Concerning imitation, there is some evidence that what has been proposed by some to be a largely automatic tendency to imitate (governed by the mirror system; see, e.g., Pickering & Garrod, 2004) is routinely mediated by executive control, so that people can avoid imitating others when such behavior might be costly or inappropriate. Imitative finger gestures are actually initiated more quickly when working memory load is *increased* (with a two-back task) than without such load (Van Leeuwen, van Baaren, Martin, Dijksterhuis, & Bekkering, 2009), suggesting that executive control is the rule (for restraining this sort of imitation from the start) rather than the exception (for adjusting this behavior later in planning).

More evidence for the importance of executive control in suppressing egocentric behavior is implicated by Brown-Schmidt's (2009) visual worlds eye-tracking study of communication. To test the role of executive control, individual differences were first measured using a Stroop task. Then subjects interacted with a confederate partner to do a referential communication task that included both shared and privileged information; subjects had to differentiate what they knew from what the partner knew. Interaction was mostly unscripted, with the confederate partner asking the subject for information using expressions that were temporarily ambiguous between an object they could both see and one that only the subject could see. Some of the time immediately after the partner asked for information, their display would disappear so that the task would be interrupted before the subject could respond (thus interrupting the grounding process), and 2 s later, the display would reappear and task would resume again. This innovative manipulation aimed to test whether subjects closely monitored the grounding process in order to keep track of what their partners were actually likely to know. The findings were clear: Subjects who were better at suppressing Stroop interference were better able to restrict themselves to considering shared (rather than private) information in the early moments of responding to their partner's temporarily ambiguous questions. And they

[5] Vogeley et al. (2001) also found ToM activation in the left temporopolar cortex.

were better able to keep track of which expressions had been verbally grounded (and could therefore be assumed to be in common ground) as opposed to which had been uttered but interrupted before being grounded (these were treated as referring to information that was still private). This is a remarkable demonstration of not only the role of executive control in perspective taking, but also the ability of interlocutors to keep detailed track of the mutual knowledge product resulting from the grounding process.

If these kinds of interactive tasks could be probed with imaging, the workings of the ToM network might be further clarified. It may be possible to use imaging to delineate a role for the mentalizing system in influencing executive control over other neural circuits (including those associated with the mirror system). Such findings would be consistent with the choice and timing evidence from our and Brown-Schmidt's experiments and could provide a mechanism by which partner-adapted information that has already been perceived or computed could have an early impact.

6.4. Mentalizing Versus Mirroring

Recall that the goal of this review is to better understand how speakers and addressees take one another into account during processing. The behavioral evidence of adaptive processing that we wish to explain emerges from not only cues that unfold during interaction (locally driven) but also simple models of a partner (globally driven; see Section 5.1). This distinction can be mapped onto its neural counterpart, the mirror system (driven by sensorimotor resonance by which one partner simulates another's perspective) versus the mentalizing system (which involves more conceptual perspective taking). How might the mentalizing and mirroring systems work together to support flexible partner-adapted processing?

The answer is not clear. In a comprehensive meta-analysis of over 200 fMRI studies, Van Overwalle & Baetens (2009) considered three possibilities: (1) that mentalizing and mirroring might show anatomical overlap and share a functional core, (2) that they might not overlap but both be active during the same sorts of tasks, or (3) that they might be activated independently. They found the mirroring and mentalizing systems to be "rarely concurrently active" (p. 564), and so concluded that they are complementary, with neither subserving the other. This conclusion does not seem like the end of the story, however. These authors acknowledge "the lack of clear anatomical definitions for the pSTS and the TPJ" and warn that the overlap in their patterns of activation "cautions against making any strong distinction between them" (p. 568). Recall that the TPJ is implicated in rapid mentalizing.

However, the seeds of an answer may exist in Noordzij et al.'s (2009) study, which aimed to distinguish mentalizing from mirror networks. Here, the right pSTS was activated not only in recognizing communicative actions, but also in planning actions intended to be recognized as communicative (see Section 6.2.2). The right pSTS, traditionally associated with the mirror system, appeared to participate in a ToM pattern of activation that included mPFC activation, as well as coinciding with the *deactivation* of the mirror system's sensorimotor areas (which were most deactivated during planning communicative action). Unfortunately this study is too new to have been covered in the meta-analysis; however, it causes us to question Van Overwalle & Baetens' conclusion that mirroring and mentalizing are independent for two reasons. First, it may have been premature to conclude that pSTS activation is indicative only of mirroring and not of mentalizing (especially given Van Overwalle & Baetens own caveat), and so the two systems may share a functional core after all. Second, it is probable that few if any studies in the meta-analysis involved interactive communication between partners (the analysis did not include the other studies deploying interactive tasks that we have surveyed here: Krach et al., 2009; Rilling et al., 2004). So it may be that deploying measures that preserve essential aspects of communicative interaction (e.g., Suda et al., 2010) along with tasks that evoke recognition and planning of communicative intentions could show more clearly how these two essential networks might work together.

6.5. Cues Hypothesized to Support Partner-Adapted Processing

In this section, we consider several cues relevant to spoken communication. As we have argued from eye-tracking and other behavioral evidence (e.g., Brennan & Hanna, 2009; Metzing & Brennan, 2003), partner-adapted processing can be both rapid and flexible. Thus it makes sense to investigate not only mentalizing as a facilitator of such behaviors, but also the role of cues or local signals about a partner's needs. Just as affordances in the environment appear to directly support behavior (Gibson, 1977; Norman, 2002), the evidence that unfolds either as feedback from a partner or progress in a joint task could shape an individual's behavior "for the partner." Reconsider (from Section 4) the three criteria that for a cue to be "communicative": it must be informative, it must be able to be perceived, and it should be able to be modified by the originator's intentions (Brennan & Williams, 1995). It can be a challenge to set up behavioral studies of communication that satisfy the last criterion. The "tacit communication game" of Noordzij et al. (2009) accomplished this quite well and found a clear dissociation for moves that signal intention versus (instrumental) moves that do not, when the moves employ otherwise identical perceptual/motor actions. As the neural network(s) associated with processing communicative intentions (whether from local cues or global knowledge) become more well understood, imaging may be able to illuminate communicative processing in ways that are impossible with behavioral studies alone.

6.5.1. Processing Cues That Initiate Social Interaction

A dialog begins when one partner recognizes another's intention to communicate. Calling a partner's name (an auditory cue) and making eye contact (a visual cue) signal the initiation of social interaction. Both of these cues activate the mPFC (in particular, the right paracingulate cortex) and the left temporal pole of an addressee (Kampe et al., 2003), suggesting that these regions are part of a multimodal circuit that supports recognizing a partner's intention to communicate.

6.5.2. Voice Cues to Partner Identity

Because fMRI studies address anatomical localization but not event-related timing, it is particularly useful to consider electrophysiological evidence from event-related potentials (ERPs) in order to consider the time course with which partner-specific information may have an effect. New evidence from electrophysiological data demonstrates that listeners integrate the content of an utterance with stereotypic information about its speaker from the earliest moments of utterance processing (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008). In Van Berkum et al.'s study, listeners heard utterances (in Dutch) whose content was either congruent or incongruent with stereotypes evoked by the voices in which they were spoken, such as: statements odd for a child speaker but not for an adult (*Every evening I drink some wine before I go to sleep*), odd for a man but not a woman (*I recently had a check-up at the gynecologist in the hospital*), or odd for a speaker with a lower-class accent but not an upper-class one (*In my free time I enjoy listening to piano music by Chopin*). Voice-incongruent utterances evoked reliable N400 waves right from the acoustic onsets of relevant words, at the same early point in time as lexically based semantic anomalies evoke N400s when other semantic information is integrated (Van Berkum et al., 2008). It is remarkable that this incongruity effect of utterance content and speaker stereotype was cued entirely by prerecorded voices (with each presented in a block). It is certainly possible that physical copresence with an interacting speaker in dialog could yield even stronger partner-specific effects, if ERP could be used in this kind of situation.

Van Berkum and colleagues next localized this speaker-specific effect. Generally speaking, recognition of a speaker's identity and characteristics (as evident in the voice) is associated with activation in the right anterior superior temporal sulcus (STS) or temporal pole. Presumably that area provides inputs into language processing in Broca's Area (BA 44 and 45 in the inferior frontal gyrus, IFG). An fMRI study using the same stimuli as Van Berkum et al. (2008) found more activation in the left IFG (or Brodmann's Areas 45/47) as well as the right IFG (BA 47) for voice-incongruent sentences than for voice-congruent sentences (although IFG was activated for both kinds of sentences; Tesink et al., 2008). This was interpreted as reflecting effort to unify lexico-semantic information from the utterance with the world knowledge stereotype evoked by the speaker's voice. Sentences in which voice and message were coherent led to enhanced activation in the bilateral superior temporal cortex (STC, BA 22 extending into BA 41), the right lingual gyrus (BA 18), and the right posterior cingulate cortex (PCC, BA 29). These regions were construed to form a "unification network" for combining linguistic and extralinguistic information, with STC activation proposed to be specific to the congruence between voice and message (as opposed to semantic coherence in general; Tesink et al., 2008). This study did not report any activation in the ToM network.

Finally, autism is associated with (and sometimes diagnosed by) ToM deficits. In another study, Tesink and colleagues tested listeners with and without autism spectrum disorder (ASD) using the same voice-incongruent and congruent stimuli. Again, the listeners were able to detect the voice-incongruent messages, showing more activation in the right IFG (BA 47) for speaker-incongruent than congruent messages (Tesink et al., 2008 2009). However, this activation was stronger in listeners with ASD than without; their increased right hemisphere activation in this area over that of non-ASD listeners was interpreted as evidence of compensation, or more effortful processing (perhaps due to difficulty in evoking stereotypes). In addition, non-ASD listeners showed more activity than did ASD listeners in the right ventral mPFC (BA 10) and right ACC (anterior cingulate cortex, BA 24/32) regions (Tesink et al., 2009).

7. Conclusions

Psycholinguistic studies of dialog that preserve as many of the natural aspects of spontaneous interpersonal communication as possible (while at the same time achieving sufficient control) have found evidence that speakers and addressees can adapt to each other from the early moments of processing. That is, processing need not be encapsulated from relevant partner-specific information that is straightforward and known in advance. Under some circumstances, speakers can adjust immediately to their addressees' needs or perspectives, even when these are distinct from their own. The following considerations, we propose, represent useful design considerations for experimental studies that aim to uncover the cognitive and/or neural bases of language processing in communicative contexts, and in particular, partner-specific processing:

- To the extent that an experimental task affords behavioral, eye-tracking, or imaging evidence that can be measured independently from evidence in the stimulus events or transcript, this gives the experimenter a window into subjects' cognitive processing.

- The "language game" that subjects are asked to play should be well characterized and staged such that it does not exclude the behavior that it aims to study. To this end, imaging studies with tasks that require subjects to communicate should yield valid data about the kind of processing that underlies language-as-action.
- Especially useful is evidence that unfolds moment by moment and can be synchronized with events or a transcript, or that can be collected from two interacting partners and synchronized.
- To experimentally distinguish "for-the-self" from "for-the-other" processing, partners doing a joint task must (at least at some point in the task) have perspectives, needs, or knowledge states that can be operationally distinguished from each other's.
- Unless the goal is to study perspective taking under cognitive load, information about one partner's needs must be available to the other partner in a timely enough fashion to be incorporated into speech planning, articulation, or interpretation—otherwise, one cannot conclude that behavior that seems to be egocentric is actually egocentric.
- It may be useful for an experimental design to distinguish local (sensorimotor) cues from global cues that are updated less often, or at least to take this distinction into account.
- It may be useful to characterize cues as to whether they consist of signals intended to be recognized as communicative (in the Gricean sense), or whether they are simply informative. This may determine whether they activate the mentalizing system.

When thinking about how to model partner-adapted processing, it is productive to consider fMRI and electrophysiology data alongside eye-tracking and behavioral studies of communication. We anticipate that timing data from electrophysiology studies and anatomical data from imaging studies have potential to clarify process models that would otherwise be ambiguous. Each approach can shape and inform the kinds of questions that the other can ask, as well as the kinds of cognitive models that it makes sense to propose. Ultimately, plausible cognitive models must be guided by neurological constraints.

The distinction between local cues and global partner models that we have developed in our behavioral studies seems to map naturally onto the mirror system and the mentalizing network, respectively. Our findings about how local and global sources of information shape one another to achieve partner-adapted processing lead us to seek out ways in which the mirror and mentalizing systems coexist in the service of language and communicative processing. Executive control appears to play an important role in both kinds of systems: for instance, to inhibit mimicry in the mirror system when necessary, and to select, suppress, or update a global perspective, especially when more than one perspective is implicated in the context (e.g., self vs. other).

The mirror system automatically processes social cues that are sensorimotor in nature (e.g., voice, gaze, body motion, backchannels), whereas ToM underlies more conceptual modeling of a partner's perspectives, needs, and intentions. It remains to be established whether and how these circuits interact. But given the range of processes they support and the likely importance of these processes in interpersonal communication, we expect that they do interact. Previous imaging studies (e.g., as surveyed by Van Overwalle & Baetens, 2009) have failed to clearly establish *how* they may work together, but this does not mean they are independent, especially since many of the tasks currently in use (especially for ToM) are based on an impoverished notion of what constitutes dialog. Most of the tasks employed so far in ToM studies have not involved interpersonal interaction (or first- or second-person communicative intent); progress could accelerate with more sophistication in the kinds of language tasks that imagers employ. Another challenge is that sometimes it is difficult to determine exactly which anatomical areas are activated in a particular study. There is much that is unknown about the potential connectivity among regions and about the time course of their activation. And it is extremely difficult to stage an experiment in a scanner that involves speaking; perhaps, new experimental techniques will make it easier to use tasks that preserve the essence of spoken (or even face-to-face) dialog, such as near-infrared spectroscopy (Suda et al., 2010).

We also expect that new evidence from imaging studies will help to clarify how ToM and mirroring neural circuits work in concert with those traditionally associated with language, with profound implications for neural models of joint processing both *within* and *between* the minds of language users. Understanding how brain networks interact may promote a more nuanced understanding of why communication failures occur, of individual differences in perspective taking, and of the neural basis of communication deficits.

In closing, we suggest that to study language use based entirely on individual cognitive processes is to overlook a ubiquitous and astonishing human skill: the coordination of the behavior and mental states of interacting individuals. Interpersonal coordination is so pervasive that it is worthy of scientific investigation in its own right. This skill proceeds in parallel (and is closely integrated) with traditional psycholinguistic processing. For that reason, we advocate studying language processing along with interpersonal coordination in order to understand what it is that minds actually do when communicating.

ACKNOWLEDGMENTS

We thank Richard Gerrig, Arthur Aron, and Hoi-Chung Leung for their comments and the Gesture Focus Group for many helpful discussions. This material is based upon work supported by NSF under Grants IIS-0527585 and ITR-0325188. Any opinions, findings, and 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

Akmajian, A., Demers, R. A., & Harnish, R. M. (1987). *Linguistics: An introduction to language and communication* (2nd ed.). Cambridge, MA: MIT Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518.

Arnold, J. E., Hudson-Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh—you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 914–930.

Arnold, J. E., Tanenhaus, M. K., Altmann, R., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science*, *15*, 578–581.

Bard, E. G., 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. G., & Aylett, M. (2000). Accessibility, duration, and modeling the listener in spoken dialogue. In: *Proceedings of Gotalog 2000, 4th Workshop on the Semantics and Pragmatics of Dialogue,* Gotalog, Sweden: Gothenburg University.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

Basden, B. H., Basden, D. R., & Henry, S. (2000). Costs and benefits of collaborative remembering. *Applied Cognitive Psychology*, *14*, 497–507.

Bekkering, H., de Bruijn, E. R. A., Cuijpers, R. H., Newman-Norlund, R., van Schie, H. T., & Meulenbroek, R. (2009). Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science (Special Issue on Joint Action)*, *1*, 340–352.

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.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic coordination in dialogue. *Cognition*, *75*, B13–B25.

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia*, *43*, 89–98.

Brass, M., Zysset, S., & von Cramon, D. Y. (2001). The inhibition of imitative response tendencies. *NeuroImage*, *14*, 1416–1423.

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

Brennan, S. E. (1991). Conversation with and through computers. *User Modeling and User-Adapted Interaction*, *1*, 67–86.

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

Brennan, S. E. (2004). How conversation is shaped by visual and spoken evidence. In J. Trueswell & M. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (pp. 95–129). Cambridge, MA: MIT Press.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2007). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, *106*, 1465–1477.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *6*, 1482–1493.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science (Special Issue on Joint Action)*, *1*, 274–291.

Brennan, S. E., & Kipp, E. G. (1996). An addressee's knowledge affects a speaker's use of fillers in question-answering. In: *Abstracts of the Psychonomic Society, 37th Annual Meeting* (p. 24)*, Chicago, IL.*

Brennan, S. E., Kuhlen, A. K., & Ratra, B. (2010). *Audience design in answering rhetorical versus sincere questions (in preparation).*

Brennan, S. E., & Ohaeri, J. O. (1999). Why do electronic conversations seem less polite? The costs and benefits of hedging. In: *Proceedings, International Joint Conference on Work Activities, Coordination, and Collaboration (WACC'99)* (pp. 227–235), San Francisco, CA: ACM.

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.

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

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, *16*, 893–900.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*, *107*, 1122–1134.

Cahn, J. E., & Brennan, S. E. (1999). A psychological model of grounding and repair in dialog. In: *Proceedings, AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems* (pp. 25–33), North Falmouth, MA: American Association for Artificial Intelligence.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, *3*(1), 1–62.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B. G., et al. (2007). The intentional network: How the brain reads varieties of intentions. *Neuropsychologia*, *45*, 3105–3113.

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., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: APA Reprinted in R. M. Baecker (Ed.), *Groupware and computer-supported cooperative work: Assisting human–human collaboration* (pp. 222–233). San Mateo, CA: Morgan Kaufman Publishers.

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

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*, 73–111.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294.

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

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken language comprehension. *Journal of Memory and Language*, *47*, 292–314.

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. Napoli & J. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (pp. 105–129). Hillsdale, NJ: Erlbaum.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*, 176–180.

Don, A., Brennan, S., Laurel, B., & Schneiderman, B. (1992). Anthropomorphism: From Eliza to Terminator 2. In: *Proceedings, CHI'92, Human Factors in Computing Systems, Monterey, CA* (pp. 67–70).

DuBois, J. (1974). Syntax in mid-sentence. In C. Fillmore, G. Lakoff, & R. Lakoff (Eds.), *Berkeley studies in syntax and semantics*, Vol. 1 (pp. III.1–III.25). Berkeley, CA: University of California.

Eden, M. (1983). Cybernetics. In F. Machlup & U. Mansfield (Eds.), *The study of information: Interdisciplinary messages* (pp. 409–439). New York, NY: John Wiley & Sons.

Ekeocha, J. O., & Brennan, S. E. (2008). Collaborative recall in face-to-face and electronic groups. *Memory*, *16*, 245–261.

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

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

Fussell, S. R., & 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.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361–377.

Galati, A., & Brennan, S. E. (2010a). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*, 35–51.

Galati, A., & Brennan, S. E. (2010b). *Audience design in the production of gesture* (under review).

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, 7, 77–83.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*, 814–821.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*, 493–501.

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Giles, H., & Powesland, P. F. (1975). *Speech styles and social evaluation*. London: Academic 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.

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

Grice, H. P. (1957). Meaning. *Philosophical Review*, *66*, 377–388.

Grice, H. P. (1975). Logic and conversation (from the William James lectures, Harvard University, 1967). In P. Cole & J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*, 596–615.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, *28*, 105–115.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61.
Harding, C. (1982). Development of the intention to communicate. *Human Development*, *25*, 140–151.
Harris, C. B., Paterson, H. M., & Kemp, R. I. (2008). Collaborative recall and collective memory: What happens when we remember together? *Memory*, *16*, 213–230.
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208–216.
Hollingshead, A. B. (1998). Retrieval processes in transactive memory systems. *Journal of Personality and Social Psychology*, *74*, 659–671.
Horton, W. S., & Gerrig, R. J. (2002). Speaker's experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589–606.
Horton, W. S., & Gerrig, R. J. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes*, *40*, 1–35.
Horton, W. S., & Gerrig, R. J. (2005b). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.
Hwang, J., Brennan, S. E., & Huffman, M. K. (2007). How non-native speakers make phonetic adjustments to partners in dialogue. In: *Abstracts of the Psychonomic Society, 48th Annual Meeting* (p. 88), *Long Beach, CA*.
Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.
Jefferson, G. (1973). A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica*, *9*, 47–96.
Jurafsky, D. (1996). A probabilistic model of lexical and syntactic disambiguation. *Cognitive Science*, *20*, 137–194.
Kampe, K. K. W., Frith, C. D., & Frith, U. (2003). "Hey John": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing", regardless of modality. *Journal of Neuroscience*, *23*, 5258–5263.
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 bias of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7, 46–50.
Kiesler, S., & Sproull, L. (1992). Group decision making and communication technology. *Organizational Behavior and Human Decision Processes*, *52*, 96–123.
Krach, S., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., et al. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neuroscience*, *10*, 9.
Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE*, *3*, e2597.
Kraljic, T., & Brennan, S. E. (2005). Using prosody and optional words to disambiguate utterances: For the speaker or for the addressee? *Cognitive Psychology*, *50*, 194–231.

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

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

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, *56*, 436–455.

Kuhlen, A. K., & Brennan, S. E. (2008). Addressees shape speaking: When confederates may be hazardous to your data. In: *Abstracts of the Psychonomic Society, 49th Annual Meeting* (p. 6)*, Chicago, IL.*

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, (in press).

Kuhlen, A. K., Galati, A., & Brennan, S. E. (2010). *Gesturing integrates top-down and bottom-up information: Effects of speakers' expectations and addressees' feedback* (under review).

Lerner, G. H. (1996). On the "semi-permeable" character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 238–276). Cambridge, MA: Cambridge University Press.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*, 78–106.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of speech code. *Psychological Review*, *74*, 431–461.

Lieberman, P. (1963). Some effects of context on the intelligibility of hearing and deaf children's speech. *Language and Speech*, *24*, 255–264.

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

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.

MacKay, D. M. (1983). The wider scope of information theory. In F. Machlup & U. Mansfield (Eds.), *The study of information: Interdisciplinary messages* (pp. 485–492). New York, NY: John Wiley & Sons.

Matthews, D. E., Lieven, E. V. M., & Tomasello, M. (2008). What's in a manner of speaking? Children's sensitivity to partner-specific referential precedents. In: *Proceedings of the LONDIAL Workshop on the Semantics and Pragmatics of Dialog, London, UK.*

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

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects in the comprehension of referring expressions. *Journal of Memory and Language*, *49*, 201–213.

Nadig, A. S., & Sedivy, J. S. (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychological Science*, *13*, 329–336.

Neider, M. B., Chen, X., Dickinson, C. A., Brennan, S. E., & Zelinsky, G. J. (2005). Sharing eyegaze is better than speaking in a time-critical consensus task. In: *Abstracts of the Psychonomic Society, 46th Annual Meeting* (p. 72)*, Toronto, Canada.*

Newman-Norlund, R. D., van Schie, H. T., van Zuijlen, A. M. J., & Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience*, *10*, 817–818.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience*, *3*, 1–13.

Nooteboom, S. G. (1991). Perceptual goals of speech production. In: *Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence, France, August 19–24* (pp. 107–110), Vol. 1.

Norman, D. A. (2002). *The design of everyday things*. New York, NY: Basic Books.
Perryman, G. A., & Brennan, S. E. (2009). Effects of multiple speakers (copresent or not) on dialog context. In: *Abstracts of the Psychonomic Society, 50th Annual Meeting, Boston, MA* .
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 167–226.
Polichak, J. W., & Gerrig, R. J. (1998). Common ground and everyday language use: Comments on Horton and Keysar (1996). *Cognition*, *66*, 183–189.
Reddy, M. J. (1979). The conduit metaphor—A case of frame conflict in our language about language. In A. Ortony (Ed.), *Metaphor and thought* (pp. 284–297). Cambridge, UK: Cambridge University Press.
Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*, 1694–1703.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.
Samuel, S. 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.
Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in ''theory of mind'' *NeuroImage*, *19*, 1835–1842.
Schober, M. F. (1998). Conversational evidence for rethinking meaning. *Social Research*, *65*, 511–534.
Schober, M. F. (2004). Just how aligned are interlocutors' representations? *Behavioral and Brain Sciences*, *27*, 209–210.
Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211–232.
Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & Language*, *14*, 131–153.
Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Science*, *10*, 70–76.
Sebanz, N., & Knoblich, G. (2008). From mirroring to joint action. In I. Wachsmuth, M. Lenzen, & G. Knoblich (Eds.), *Embodied communication* (pp. 129–150). Oxford: Oxford University Press.
Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*, 353–367.
Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: University of Illinois 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*, 1–38.
Shockley, K., Richardson, D., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*, 305–319.
Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language*, *32*, 25–38.
Stellmann, P., & Brennan, S. E. (1993). Flexible perspective-setting in conversation. In: *Abstracts of the Psychonomic Society, 34th Annual Meeting* (p. 20), *Washington, DC*.
Suda, M., Takei, Y., Aoyama, Y., Narita, K., Sato, T., Fukuda, M., et al. (2010). Frontopolar activation during face-to-face conversation: An in situ study using near-infrared spectroscopy. *Neuropsychologia*, *48*, 441–447.
Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*, 81–94.
Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Handbook of perception and cognition, Vol. 11: Speech language and communication*. San Diego, CA: Academic Press.

Tanenhaus, M. K., & Trueswell, J. C. (2004). Eye movements as a tool for bridging the language-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-action traditions* (pp. 3–37). Cambridge, MA: MIT Press.

Tesink, C. M. J. Y., Buitelaar, J. K., Petersson, K. M., van der Gaag, R. J., Kan, C. C., Tendolkar, I., et al. (2009). Neural correlates of pragmatic language comprehension in autism spectrum disorders. *Brain*, *132*, 1941–1952.

Tesink, C. M. J. Y., Petersson, K. M., van Berkum, J. J. A., van den Brink, D., Buitelaar, J. K., & Hagoort, P. (2008). Unification of speaker and meaning in language comprehension: An fMRI study. *Journal of Cognitive Neuroscience*, *21*, 2085–2099.

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*, 580–591.

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., & Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience*, *4*, 165–184.

Van Duynslaeger, M., Van Overwalle, F., & Verstraeten, E. (2007). Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Social Cognitive and Affective Neuroscience*, *2*, 174–188.

Van Leeuwen, M. L., van Baaren, R. B., Martin, D., Dijksterhuis, A., & Bekkering, H. (2009). Executive functioning and imitation: Increasing working memory load facilitates behavioural imitation. *Neuropsychologia*, *47*, 3265–3270.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564–584.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., et al. (2001). Mind reading: Neural mechanisms of theory of mind and self perspective. *NeuroImage*, *14*, 170–181.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 1160–1175.

Wiener, N. (1965). *Cybernetics—or control and communication in the animal and the machine* (2nd ed.). Cambridge, MA: The MIT Press (originally published in 1948).

Wiley, J., & Bailey, J. (2006). Effects of collaboration and argumentation on learning from web pages. In A. M. O'Donnell, C. E. Hmelo-Silver, & G. Erkens (Eds.), *Collaborative learning, reasoning, and technology* (pp. 297–321). Hillsdale, NJ: Erlbaum.

Wiley, J., & Jensen, M. (2006). When three heads are better than two. In: *Proceedings, CogSci 2006, 28th Annual Conference of the Cognitive Science Society* (p. 3275), Vancouver, CA: Cognitive Science Society.

Wilkes-Gibbs, D. (1986). *Collaborative processes of language use in conversation*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Wright, D. B., & Klumpp, A. (2004). Collaborative inhibition is due to the product, not the process, of recalling in groups. *Psychonomic Bulletin & Review*, *11*, 1080–1083.

Yngve, V. H. (1970). On getting a word in edgewise. In: *Papers from the 6th Regional Meeting of the Chicago Linguistic Society* (pp. 567–578), Chicago, IL: Chicago Linguistic Institute.