17

Mindreading in an Exotic Case

The Normal Adult Human

DALE J. BARR
BOAZ KEYSAR

The world that is most important to humans is the social world—not the mute world of objects, but the world of living, acting minds. Dealing with other minds introduces a fundamental element of uncertainty into life, because the beliefs and desires that drive other people's behavior are hidden from view. Yet, adult humans are highly competent navigators of the social world, because a lifetime of experience with other people has made them experts at reasoning in mentalistic terms. Thus, normal adults can be said to possess a theory of how other minds work that enables them to impute motivations, detect deceptions, and more generally predict and explain the behavior of others. This so-called theory of mind serves as an important foundation for human interaction, making it possible for individuals to coordinate activities, including activities as complex as holding a conversation. An active goal of research in cognitive science is to understand the nature of human "mindreading"—not, of course, mindreading in the magical or paranormal sense but, rather, in the mundane sense of how people apply theory of mind in order to infer the mental states of other people.

Much of what we know about mindreading in humans derives from

research on cognitive development. The development of mindreading abilities throughout the lifespan can be characterized as a trajectory away from egocentrism and toward greater and more nuanced mental attribution. Although most normal functioning adults take it for granted that different people can have very different perceptions of reality, developmental research suggests that this understanding unfolds gradually. Many of the rudiments of mindreading seem to be in place well before the third year, such as the ability to discern goal-directed actions (Woodward, 1998) and to appreciate that the behaviors of others is functionally organized in terms of desires (Wellman, 1991). However, an adult-like understanding of other minds is not in place until children appreciate that others can have false beliefs about the world. This critical development is believed to take place between the ages of 4 and 6 years. Children younger than this age seem to have difficulty distinguishing what they believe from what others believe (Wimmer & Perner, 1983), a finding that has been replicated in different cultures with many different tasks (Wellman, Cross, & Watson, 2001).

Another important source of insight into human mindreading abilities comes from research on nonnormative and animal cases. The underlying nature of human mindreading abilities has been compared and contrasted with those of chimpanzees (Premack & Woodruff, 1978). Currently, controversy surrounds the question of whether these evolutionarily close relatives have a human-like theory of mind (Povinelli & Vonk, 2003; Tomasello, Call, & Hare, 2003). Researchers have also explored the mindreading and communicative abilities of brain-damaged individuals and have proposed theories regarding the neural circuits that underlie these abilities in normal populations (e.g., Sabbagh, 2004).

In much theorizing on mindreading in humans, the normal adult human appears largely as a figure in the background, an ideal against which the abilities of the chimpanzee, the child, and brain-damaged individual are compared. However, much of what is known about mindreading processes in normal human adults is indirect and based on extrapolation from disordered or child populations—not to mention a lot of common sense. In this chapter, we address the question: What can be learned from directly studying theory-of-mind processes in adults?

To provide an overview, we discuss findings from a research program investigating an important aspect of adult mindreading—a listener's reasoning about the beliefs of a speaker in order to resolve ambiguity in conversation. This research program has revealed some unexpected limits on this aspect of adult mindreading. In particular, our investigations indicate a large degree of egocentrism in how normal adult listeners interpret a speaker's utterances. Across a range of experiments and tasks, adult listeners do not appear to reliably take into account the knowledge

that they share with the speaker when they interpret the meaning of what that speaker has said. Given the assumption that communication is typically successful, this seems to indicate that mindreading might play less of a role in certain aspects of language processing than is generally assumed. We suggest that reasoning about a speaker's beliefs occurs primarily as part of self-monitoring, a process through which listeners monitor their interpretations for errors and correct them when they are detected.

Our findings also suggest that the egocentrism observed in 3-year-old children and in individuals with frontal brain damage is also present in normal adults. By using eyetracking techniques, we have been able to observe the process of interpreting another's actions as it unfolds through time. This online methodology suggests that the egocentrism of the young child does not disappear, but lives on in the early moments of adult processing. Although the end product of adult processing shows more sensitivity to others' beliefs, this is not because adults are less likely than children to initially interpret a speaker's action egocentrically. Rather, it is because they are more likely to detect and correct interpretive errors before they *act* upon them. Thus, our data suggest an underappreciated continuity in social reasoning between young children and normal adults.

For the sake of clarity, we begin by noting that the term "mindreading" has been used rather liberally to denote a broad variety of activities related to social cognition. As Ames (2004; Chapter 10, this volume) makes clear, mindreading can involve the application of abstract schematic knowledge or stereotypes, the projection of one's own beliefs or desires onto a target individual, or active perspective taking. In addition, there are marginal cases such as automatic mimicry and emotional contagion (see Hodges & Wegner, 1997, for a review), which may or may not be considered a form of mindreading. One factor that varies across these activities is the extent to which they require the use of metarepresentations, that is, representations about another person's representations. In this chapter, what we are interested in is how people use metarepresentations—specifically, their beliefs about another person's beliefs—in interpreting the actions or utterances of that person.

There is an important distinction between the processes underlying the use of metarepresentations versus the application of stereotypic knowledge structures such as schemata or scripts in order to predict another's behavior. When Albert's roommate Brenda sees him get up from his chair in the living room and walk toward the refrigerator, she might recognize this pattern of behavior as an instance of a well-known script, "going to the fridge." Activation of this script might lead her to impute to Albert the goal of getting something to eat or drink. Although the end

result might be a metarepresentation about what Albert desires, the process by which this inference was drawn may not have required metarepresentation. As work by Read and Miller (1998) has shown, this form of inference can be approximated by a neural network. The architecture of their network is based upon the interactive activation network that McClelland and Rumelhart (1981) used to simulate the perception of written words. Thus, the mechanisms underlying schema- or script-based inferences about goals might be identical to the domain-general pattern recognition mechanisms involved in perception. However, if Brenda has reason to believe that Albert thinks the fridge is empty, then the task of imputing a motive to Albert shifts from one of pattern recognition to one of problem solving and decision making. Under such circumstances, Brenda must try to see the world through Albert's eyes in order to predict and explain his behavior, because there is no prior pattern to which the behavior can be matched. She must attempt to retrieve relevant knowledge about Albert, such as the fact that he wants to go downtown, and is walking toward the fridge in order to consult the bus schedule that is attached to its door. In contrast to the pattern recognition case, Brenda's interpretation of Albert's behavior in this situation would seem to involve particularized inferences about Albert's current state of mind. It is this latter sense of mindreading that we address in this chapter.

Studies of how people draw and use metarepresentational inferences of this sort have typically focused on children. However, the normal adult has not been entirely neglected as an object of research. For example, researchers have addressed the neural substrates of theory of mind using normal adults (for a review, see Siegal & Varley, 2002). Another area in which adult perspective taking has been investigated is social attribution (for a review, see Nickerson, 1999). This research has been focused on questions such as how people estimate how they are seen by others (Gilovich, Savitsky, & Medvec, 1998), how people reason about others' construal of a prevailing situation (Griffin, Dunning, & Ross, 1990), or how people impute knowledge to people in a particular social group (Fussell & Krauss, 1991). Much of this work suggests that the outcome of the social reasoning processes is egocentrically biased, with people assuming that other people know the things that they know (Nickerson, 1999).

Although these studies continue to provide valuable insights into social judgment, the cognitive mechanisms by which people put such judgments to use have not been fully specified. What is lacking is a detailed processing model of theory of mind and an understanding of how this system interfaces with other cognitive systems, such as the language processing system. Experimental studies with normal adults can play a

critical role in the development of such a model, because they allow for clear isolation of the individual factors that might be involved in mindreading.

Our own interest in the metarepresentational aspect of mindreading derives from our research on how people resolve ambiguity in language comprehension. Theories of language use have long placed metarepresentations at the heart of communication (Clark & Marshall, 1981; Grice, 1957). A basic tenet of modern theories of language use is that language is inherently ambiguous—the same utterance can mean different things, depending upon the speaker's intention. Speakers and listeners can reduce this ambiguity by processing utterances against the background of mutual knowledge, or *common ground*—the set of information that is shared, and critically, *known to be shared* (Clark & Carlson, 1981; Clark & Marshall, 1981). This emphasis on shared knowledge implies that language users maintain and routinely consult particularized models of what their interlocutors know when they process language. Indeed, Clark and Marshall (1981) cogently argue that processing utterances against common ground is the only true guarantee that a communicative act will succeed. Given the intuition that language users are routinely successful at achieving shared understanding, one might expect to find that metarepresentations strongly constrain how normal adults process language.

We began our research by asking the following question: How does a listener's knowledge about what the speaker knows affect how the listener interprets what the speaker says? To answer this question, we have used eye-tracking techniques to monitor the listener's comprehension process (e.g., Keysar, Barr, Balin, & Brauner, 2000). In these studies, participants played the role of "listener" in a communication game with a confederate, who played the role of "speaker." The speaker and listener worked together to rearrange a set of objects in a mutually visible, vertical set of shelves, or "grid," that the experimenter placed between them (see Figure 17.1). The speaker received a diagram showing how the objects should be positioned in this grid. The task was for the listener to follow the speaker's instructions to move objects from slot to slot in order to match the diagram. We occluded the contents of certain slots in the grid so only the listener, but not the speaker, could view them. This created a difference in perspective between the interlocutors, such that only a subset of the objects that the listener knew about was also known to the speaker. For instance, one of our grids contained a large and a medium-sized candle that were mutually visible to both the speaker and the listener. This same grid also had a slot that was occluded from the speaker's view, which contained an even smaller candle that was visible only to the listener. At a certain point the speaker delivered an instruc-

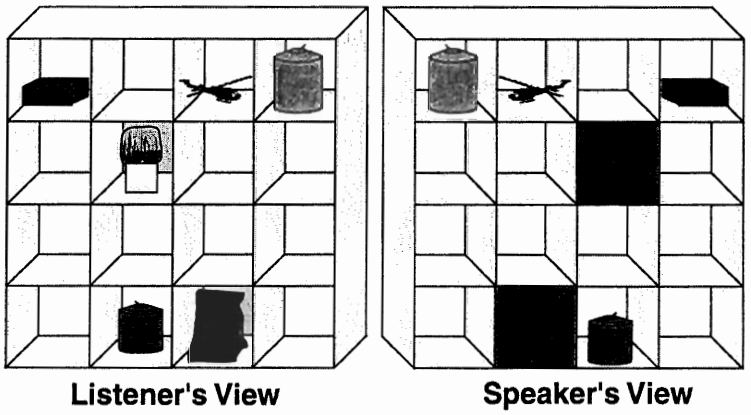

FIGURE 17.1. A grid used in the study from the listener and the speaker's view. The competitor (small candle) or control object is hidden in the bag in the bottom row.

tion that we call the "critical instruction." For example, the speaker instructed the listener to move the "small candle." Note that from the listener's perspective the best match for this expression is the smallest of the three candles. However, the listener does not have any reason to believe that the speaker knows of the existence of this smallest candle. Therefore, he or she should identify the medium-sized candle as the one intended by the speaker (henceforth the "target" object), because that is the smaller of the two candles that the speaker knows about. The listener should ignore this smaller, hidden candle (the "competitor"), even though it is a better match to the speaker's expression, because it is not in common ground. To assess the extent to which listeners still considered the competitor object, we contrasted this test condition with a control condition in which the competitor was replaced with an object that did not match the speaker's expression (e.g., a toy monkey). Our question was how effectively listeners could make use of their common ground with the speaker in order to identify the target.

Our initial studies using this paradigm revealed that listeners strongly considered competitors as referents for the speaker's expressions. For example, they spent more time looking at the competitor (small candle) than the control object (toy monkey), even though they knew the speaker was ignorant of the identity of hidden objects (see Keysar et al., 2000, for details; Keysar & Barr, 2002, for a review of related findings). Occasionally (in about 20% of cases), listeners even attempted to pick up and move competitor objects, although they tended to eventually select the target. These findings suggest a surprising

amount of egocentrism in how normal adults interpret referential expressions.

However, do these findings reflect a genuine failure to consider the speaker's perspective, or is there some more compelling alternative explanation? We conducted a set of studies intended to rule out certain alternative explanations and test the robustness of this egocentrism (for a full report, see Keysar, Lin, & Barr, 2003). For example, one such explanation is that perhaps listeners are not egocentric per se, but have difficulty ignoring competitors that are, so to speak, staring them right in the face. More commonly in real-world communication, the non-common ground "competitors" include things that the listener knows about but that are not perceptually available at the time of speaking. Thus, in the experiment described below we sought to eliminate the perceptual availability of competitors by having listeners hide them not only from the speaker but also *from themselves* by placing these objects inside a brown paper bag. The speaker could hear the rustling of the bag, but because of a visual barrier she could not witness the listener hiding the object. Thus, the listener would know of the existence of the competitor, but would not be able to see it at the moment of the critical instruction. If the listener still considered the competitor, this would suggest that the effect is not due to some sort of low-level perceptual interference, but is due to the listener's egocentrism.

We also sought to sharpen the listener's awareness that the speaker's perspective was different by including a condition in which listeners were led to believe that the speaker was not ignorant of the true identity of the occluded object but, instead, had a false belief about it. For example, after the experimenter gave the listener a small candle to hide in the bag, she ostensibly misled the speaker regarding its identity by showing her a picture of a different object (e.g., a small plastic truck) and telling her that this was the object that the listener was hiding. The listener saw the picture and witnessed the experimenter showing it to the confederate. Thus, the listener would be led to believe that the speaker thought that the object in the bag was a truck when it really was a small candle. Listeners were in collusion with the experimenter, who had secretly instructed them beforehand that she would occasionally mislead the speaker about the identity of the object in the bag, and requested that they not reveal its true identity. We contrasted this *false-belief* condition with a condition similar to our previous studies, in which the speaker was not provided any information about the identity of the hidden object (*ignorance condition*).

Would listeners still consider competitors as referents for the speaker's expressions even though they were no longer perceptually available? Furthermore, would they still do so even when they believed that the

speaker had a false belief about the identity of the hidden object? We assessed the consideration of hidden objects by examining various eye-tracking measures from the onset of the critical phrase (e.g., "the small candle") up to the moment at which listeners selected the target object. As in previous studies (e.g., Keysar et al., 2000), the pattern of results revealed a strong degree of egocentrism in our normal adult participants. Even though the competitors were not perceptually available (e.g., the candle was in a bag), they still caused substantial interference. When the occluded slot contained a competitor (e.g., the smallest of the three candles), listeners fixated the occluded slot four times as often and for over three times as long as they did when it contained a control object (e.g., the monkey). They also took significantly longer to notice the intended referent (e.g., the smaller of the two larger candles) when the bag contained a competitor.

Most surprisingly, in about 25% of the trials in which the bag contained a competitor, listeners behaved completely egocentrically, picking up the bag and moving it instead of the target object, and had to be corrected by the confederate. Viewed more optimistically, that means that 75% of the time, listeners spontaneously chose the correct target object. However, such optimism must be tempered by the fact that a full 71% of participants moved a competitor on at least one out of the four trials containing a competitor. After listeners realized their blunder, they became much less likely to do it a second time. Moreover, it is important to consider not only the end product of interpretation but also the process that yielded that interpretation. Even though 71% of our listeners made at least one error, nearly all (95%) were delayed in selecting the target when there was a competitor, regardless of whether they moved it or even looked at it. Thus, evidence for egocentrism was present in the eye movement data even when it would not have been revealed by the listener's final judgment.

Even more surprising was that listeners appeared to experience just as much interference from the competitor when they thought that the speaker was *misinformed* about the identity of the competitor as when they thought he or she was ignorant. In other words, even when the listener thought that the speaker believed that the hidden small candle was really a toy truck, upon receiving the instruction to "move the small candle," they looked at the hidden bag containing the candle just as much as they did when they thought that the speaker was ignorant about the contents of the bag. They were also just as likely to move the competitor in the false-belief condition as in the ignorance condition. Relative to the control condition, they were also equally slow to realize that the smaller of the two mutually available candles was the only plausible referent from the speaker's perspective.

These findings highlight the point that even when listeners are clear about what the speaker believes, they have trouble putting this information to use in interpreting the meanings of their actions. To account for such effects, we have put forth the perspective adjustment model (Keysar & Barr, 2002; Keysar, Barr, & Horton, 1998). Perspective adjustment is an anchoring and adjustment model in the spirit of the decision model of Tversky and Kahneman (1974). According to perspective adjustment, listeners initially "anchor" their interpretations of expressions in available information without regard to the other's perspective. By using common ground, listeners can adjust toward the speaker's perspective, although this adjustment is optional—and typically insufficient. This means that there is a systematic source of misunderstanding in language comprehension—namely, a listener's failure to sufficiently discount information not known to the speaker. (For a related model, see Epley, Keysar, Van Boven, & Gilovich, 2004.) As the structure of this model makes clear, we believe that people do make mentalistic inferences during conversation; however, such inferences are not automatic or obligatory, and may have a limited role in guiding the earliest moments of comprehension.

The possibility that inferences about common ground are optional, or are made largely as a kind of afterthought, means that there is no guarantee that communication will be successful in any given case. However, this might not be such a bad thing if what is typically available to the listener also tends to be available to the speaker. If this is the case, then access to a speaker's beliefs may not be a necessary prerequisite to successful communication. Such an idea is admittedly controversial, given the avowed ambiguity of communication (although see Recanati, 2002, for a philosophical defense of this position). As opposed to the potential ambiguity that is latent in any given utterance, the *actual* ambiguity that language users experience is a function of the degree of alignment between their perspectives. At this point, unfortunately, we know little about the factors that cause perspectives to diverge or converge, because this is still an emerging line of research. Existing research does, however, claim that the perspectives of a speaker and a listener may come into alignment through low-level, resource-free implicit mechanisms such as priming and associative learning (Barr, 1999; Barr & Keysar, 2002; Garrod & Anderson, 1987; Garrod & Doherty, 1994; Markman & Makin, 1998; Pickering & Garrod, ~~in press~~ 2004). Although processes such as associative learning or priming might seem too simplistic to hold things together during a conversation, recent research using multiagent computer simulation indicates that such mechanisms may be sufficient to support coordinated communication in a community where mindreading is absent (Barr, 2004; Steels, 1998).

Furthermore, the egocentrism of language users makes sense when assessed against the feedback-rich environment of situated language use (Barr & Keysar, 2004). Because language users consistently monitor and provide feedback to their interlocutors (Clark & Brennan, 1991; Clark & Krych, 2004), there are many opportunities in natural dialogue to interactively diagnose and correct miscoordination. Of course, how people actually use this feedback to improve coordination is likely to involve access to common ground, which is precisely the role that the perspective adjustment model stipulates for this kind of knowledge. We hope that this observation drives home the point that our argument is not about whether people ever draw metarepresentational inferences during conversation—clearly, they do—but concerns *when* they do and *how* they put these inferences to use.

Our findings, which suggest that mental state attribution may not be fully integrated into language processing, are consistent with the observation that children become sophisticated language users before they become sophisticated mindreaders. Thus, a child 4 years old would fail the false-belief task would have little trouble understanding and producing references in speech, even though the metarepresentational capacity has been viewed as a *sine qua non* of successful reference (e.g., Clark & Marshall, 1981). In fact, Epley, Morewedge, and Keysar (2004) demonstrate that adults are different from children not in the initial egocentric process but in their ability to effectively recover from an error. Using a paradigm similar to the one we described, Epley and colleagues (2004) show that adults look at the hidden competitor just as quickly as children, but they are faster to identify the target and less likely to move the competitor. This demonstrates that the early moments of comprehension are the same for children and adults. However, a critical difference between the adult and child is the adult's ability to self-monitor and preempt or recover from an egocentric error.

The study of mindreading in normal adults suggests a continuity in processing not only between children and adults but perhaps even between the normal case and people with prefrontal brain damage. Research on such patients suggests that these individuals have difficulty in inhibiting prepotent responses that are cued by environmental stimuli but irrelevant to their current goals (see Miller & Cohen, 2001, for a review). For example, individuals with prefrontal damage produce more errors on the Stroop task than normals (Vendrell et al., 1995). We have found a task that, in essence, can cause normal individuals to occasionally behave in ways that are similar to such patients, in the sense that listeners have difficulty inhibiting the selection of the competitor, which was always a better referent for the speaker's expression. An interesting prediction from our study is that if adults are placed under a severe cog-

nitive load that would inhibit their ability to self-monitor, the difference between normal adults and frontal patients might be diminished. The load manipulation might have little effect, however, on the speed of performance in the control condition, suggesting that the core language abilities would remain intact. Finally, it is an interesting question whether adults under cognitive load might perform like children on false-belief tasks or other tasks involving mental attribution. This result might be expected, considering that cognitive load has been shown to interfere with perspective-taking processes (Hodges & Wegner, 1997).

To conclude, we hope to have presented some compelling arguments why adult theory of mind should be an object of investigation in its own right. Research has only scratched the surface of the complexities of the cognitive mechanisms underlying mindreading in the normal adult. Adults represent the endpoint of development, and as such they provide a context for understanding developments in the young child. Likewise, to understand what mechanisms are absent or impaired in the case of people with mindreading deficits, such as in the case of prefrontal damage, it is important to have a standard of comparison that is empirically grounded. Currently, much research tacitly assumes an adult mindreading competency that is sophisticated, routinely accessed, and tightly integrated with other cognitive functions. Against the background of such an ideal, the egocentric behavior of the normal adults that we have observed in our laboratory appears quite exotic indeed.

REFERENCES

Ames, D. R. (2004). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology, 87*, 340–353.

Barr, D. J. (1999). *A theory of dynamic coordination for conversational interaction.* Unpublished PhD thesis, The University of Chicago.

Barr, D. J. (2004). Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science, 28*, 937–962.

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

Barr, D. J., & Keysar, B. (2004). Making sense of how we make sense: The paradox of egocentrism in language use. In H. L. Colston & A. N. Katz (Eds.), *Figurative language comprehension: Social and cultural influences.* Mahwaw, NJ: Erlbaum.

~~Barr, D. J., & Keysar, B. *Misunderstanding in language comprehension: Learning to monitor common ground.* Unpublished manuscript.~~

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared*

delete: not cited in text.

no: Year?

cognition (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 313–330). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81.

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

Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults: Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology, 40*, 760–768.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*, 327–339.

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

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

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

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346.

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

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality & Social Psychology, 59*, 1128–1139.

Hodges, S. D., & Wegner, D. M. (1997). Automatic and controlled empathy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 311–339). New York: Guilford Press.

Keysar, B., & Barr, D. J. (2002). Self-anchoring in conversation: Why language users don't do what they "should." In T. Gilovich, D. W. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 150–166). Cambridge, UK: Cambridge University Press.

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

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.

Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General, 127*, 331–354.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of

context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375–405.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience, 24*, 167–202.

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737–759.

Pickering, M. J., & Garrod, S. (~~in press~~ 2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169-226.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in cognitive sciences, 7*, 157–160.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515–526.

Read, S. J., & Miller, L. C. (1998). On the dynamic construction of meaning: An interactive activation and competition model of social perception. In S. J. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 27–70). Mahwah, NJ: Erlbaum.

Recanati, F. (2002). Does linguistic communication rest on inference? *Mind and Language, 17*, 105–126.

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition, 55*, 209–219.

Siegal, M., & Varley, R. (2002). Neural systems involved in "theory of mind." *Nature Reviews Neuroscience, 3*, 463–471.

Steels, L. (1998). Synthesizing the origins of language and meaning using coevolution, self-organization and level formation. In J. R. Hurford, M. Studdert-Kennedy, & C. Knight (Eds.), *Approaches to the evolution of language: Social and cognitive bases* (pp. 384–404). Cambridge, UK: Cambridge University Press.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—the question is which ones and to what extent. *Trends in cognitive sciences, 7*, 153–156.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Vendrell, P., Junque, C., Pujol, J., Jurado, M. A., Molet, J., & Grafman, J. (1995). The role of prefrontal regions in the Stroop task. *Neuropsychologia, 33*, 341–352.

Wellman, H. M. (1991). From desires to beliefs: Acquisition of a theory of mind. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 19–38). Oxford, UK: Blackwell.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684.

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.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.