Research Article

TAKING PERSPECTIVE IN CONVERSATION:
The Role of Mutual Knowledge in Comprehension

Boaz Keysar, Dale J. Barr, Jennifer A. Balin, and Jason S. Brauner

The University of Chicago

Abstract—*When people interpret language, they can reduce the ambiguity of linguistic expressions by using information about perspective: the speaker's, their own, or a shared perspective. In order to investigate the mental processes that underlie such perspective taking, we tracked people's eye movements while they were following instructions to manipulate objects. The eye fixation data in two experiments demonstrate that people do not restrict the search for referents to mutually known objects. Eye movements indicated that addressees considered objects as potential referents even when the speaker could not see those objects, requiring addressees to use mutual knowledge to correct their interpretation. Thus, people occasionally use an egocentric heuristic when they comprehend. We argue that this egocentric heuristic is successful in reducing ambiguity, though it could lead to a systematic error.*

Language is inherently ambiguous—every linguistic expression can convey more than one intention. Comprehension requires ambiguity resolution at all levels of linguistic processing, from the perception of phonemes, through syntactic parsing, to the identification of the speaker's intention. Addressees use a variety of sources of information to reduce the ambiguity of linguistic expressions. For example, people use visual context to restrict the range of possible syntactic structures a sentence might have (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), and they use context to constrain lexical ambiguity (e.g., Glucksberg, Kreuz, & Rho, 1986). One source of information that could be important in resolving ambiguity is knowledge regarding who knows what. We investigated the role such knowledge about perspective plays in ambiguity resolution.

Traditional theories of language use suggest that addressees routinely use shared or mutual perspective in comprehension (e.g., Clark & Carlson, 1981; Clark & Marshall, 1981; Clark, Schreuder, & Buttrick, 1983). These theories assume that comprehension is restricted to mutually known information. Such theories typically take a macro-level approach, describing only the product of comprehension. Our research takes a more fine-grained, micro-level approach to identify the mental processes that underlie perspective taking in comprehension. Specifically, we studied such processes during referential communication, in which speakers refer to objects and addressees attempt to identify these objects.

Imagine the following situation. Your friend Jill comes into your office, sits down on the other side of your desk, and says, "Can I see this book? It looks very interesting." When she says "this book," she is referring to *The Social Life of the Yellow Slug*, by H. Dull, a book that is on the desk between the two of you. But how would you know which book Jill is referring to? After all, your shelves are filled with many books, and each one of them is a perfect referent for the description "this book."

You might immediately identify the intended book because you rely on common ground with your friend. In this case, the book is in common ground because it is physically or perceptually co-present. You identify it by using what Clark and Marshall (1981) called a co-presence heuristic. Given that you can mutually see the book, you conclude that it is the intended referent. You resolve ambiguity by restricting the set of potential referents to mutual knowledge with your friend.

But consider an alternative comprehension strategy that people might use in conversation, an egocentric heuristic. We define an egocentric heuristic as a tendency to consider as potential referents objects that are not in common ground, but are potential referents from one's own perspective.[1] Suppose that in addition to the book on yellow slugs, you have on your desk a book that you really liked, *The Sex Life of Spiders*, by D. Excite. But Jill cannot see the book on spiders because your computer monitor is in the way. Clearly, this book is not part of your common ground with Jill, but to the extent that you do use an egocentric comprehension strategy, you might consider it as a referent.

If you select the book about the spiders as the referent, you make an error. Indeed, you might pick it up only to be corrected by Jill, who says that she was actually talking about the book about the slugs. But most likely, you would use your knowledge about common ground to correct your interpretation before even reaching for the wrong book. After all, you know that she cannot see the spider book, and therefore she could not have referred to it.

We therefore propose that (a) addressees occasionally use an egocentric strategy, considering potential referents even when they know that these referents are inaccessible to the speaker, and (b) mutual knowledge is used to correct interpretation errors that result from such an egocentric interpretation. We report here two experiments that investigated our proposal in the context of a real conversation. When addressees use an egocentric strategy, and they consider referents that are invisible to the speaker, it should interfere with their ability to detect the shared, intended referent. We show that addressees occasionally consider the hidden spider book and that they consequently take longer to identify the shared referent, the slug book, because they need to correct their initial error. In these cases, they use mutual knowledge as error correction. Moreover, we show that sometimes the egocentric interpretation is so compelling that addressees make an error—they pick up the spider book instead of the slug book.

EXPERIMENT 1

In the first experiment, participants played a version of the referential communication game (e.g., Glucksberg, Krauss, & Higgins,

---

Address correspondence to Boaz Keysar, University of Chicago, Department of Psychology, 5848 South University Ave., Chicago, IL 60637; e-mail: boaz@ccp.uchicago.edu.

1. By "egocentric," we do not mean that your search for referents is insensitive to mutual knowledge, but that you would consider referents even though they are not in common ground; your search might still be attenuated by the fact that these objects are not shared with the speaker.

1975; Krauss & Glucksberg, 1977; Krauss & Weinheimer, 1964). A participant and a confederate sat on the two sides of a vertical array of slots with objects (see Fig. 1). The confederate, the "director," received a picture that displayed the objects in a different organization. The director instructed the participant, the "addressee," to move objects around so that they would be in the same positions as objects in that picture. For example, if the director's photograph showed the truck from Figure 1 in the slot below its initial position, then the director might say, "Put the truck one slot down."

We wanted to be able to tell which objects addressees would consider as referents of expressions such as "the truck," as they heard the instructions. A recent application of an eye-tracking methodology has introduced a novel way to experimentally investigate comprehension processes in real time (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Tanenhaus et al., 1995). Tanenhaus and his colleagues have successfully used eye fixation data to investigate the way information is integrated during language understanding. When people identify a referent, their eyes quickly fixate on that object, as a precursor to a reach toward it. Eye fixation, then, is a sensitive measure that can index comprehension in time. We collected eye fixation data from addressees to evaluate which objects they considered as referents.

To study the role of perspective in the comprehension process, we distinguished between the perspectives of the director and the addressee by blocking the contents of some of the slots from the director's view (see Keysar, 1997, for a methodological discussion). For example, in Figure 1, all three candles were visible to the addressee, but the director could see only two of them. The smallest candle was occluded from the director's view, so he could not know about it. This perspective difference provided a critical test of the hypothesis. Suppose the director said, "Now put the small candle above it." Clearly, the director would be referring to the small candle that he could see. However, if the addressee used an egocentric interpretation strategy, then he or she would initially consider the occluded candle as the intended referent.

Method

Participants

Twenty native English speakers participated in the experiment as addressees.

Experimental setup

An array of 4×4 slots contained the objects on a table between the director and the addressee. We used an Applied Science Laboratories eyetracker to follow the addressee's eye movements. The addressee wore a headband with a small camera lens that filmed the left eye, and a magnetic head tracker provided information about head movement in space. Together, this information determined the addressee's direction of gaze. A freestanding video camera filmed the grid of objects on HI8 video, and the participants' conversation was recorded on the same videotape. We had two sources of data about the addressee's eye position: a superimposed crosshair on a video image of the grid from the addressee's point of view (sampled at 30 Hz) and a computer
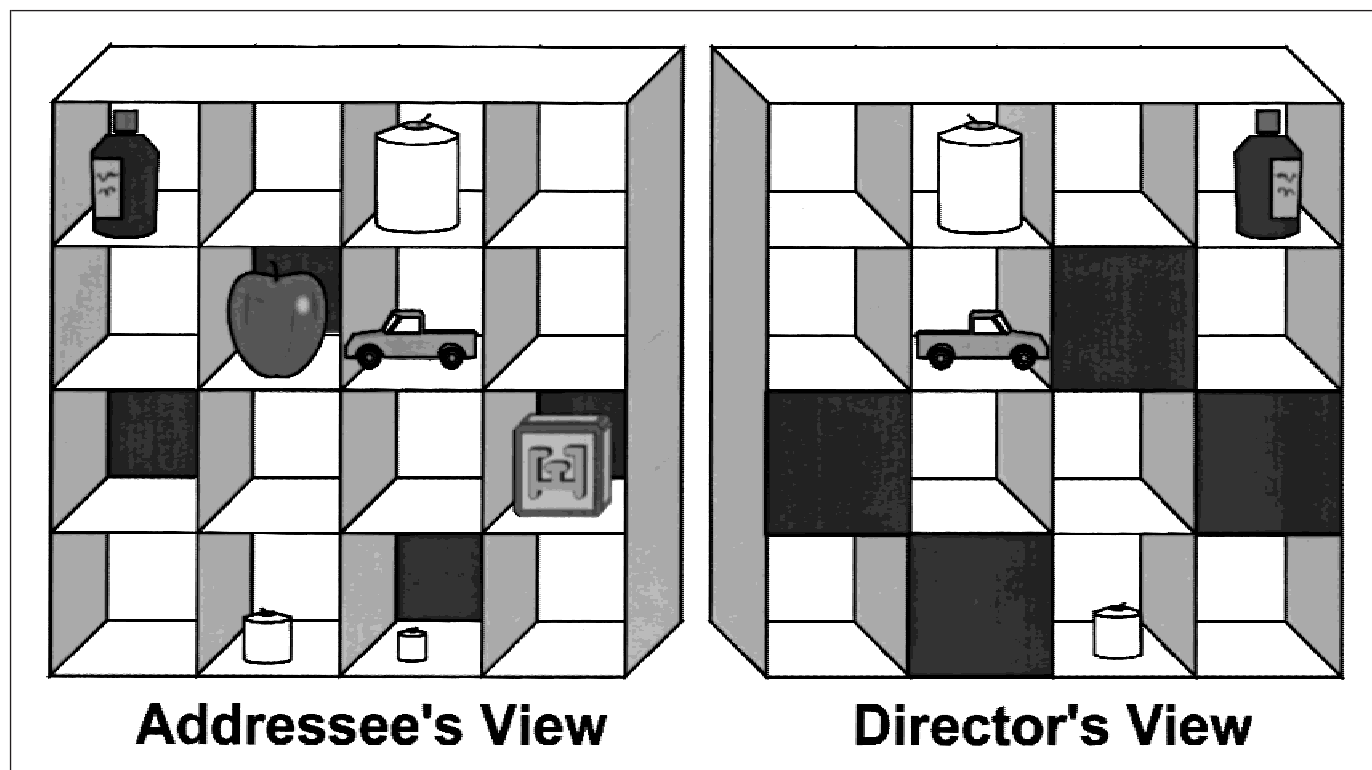
Fig. 1. The 16 slots with a typical set of objects. The addressee's and director's views are distinct because of the occluded slots. The critical instruction (referring to "the small candle") picks out a different candle from the director's perspective (shared candle) than from the addressee's perspective (occluded candle).

file that digitally stored the spatial coordinates (sampled at 60 Hz). Addressees were free to move their heads, allowing fairly natural interaction.

Procedure

A male undergraduate research assistant was the director, to ensure uniformity across participants. Addressees were led to believe that the roles of director and addressee were assigned randomly. The experimenter then explained that they would be playing a communication game, and they would need to reorganize the objects in the grid in each trial. The director received a photograph of the grid in which the objects had the desired final location, and then guided the addressee in moving objects around to match the photograph. The slots the director could not see were similarly blocked in the photograph, so it was clear that only the mutually visible objects were part of the game. Unbeknownst to the addressee, the director followed a script. At some point in the trial, the director would give the critical instruction, which was always ambiguous: The shared perspective yielded one referent, and the addressee's perspective yielded another. To make sure the addressee fully understood the director's perspective, the players switched roles for one trial during practice.

To provide a motivation for this simple task, the experimenter explained that we made it more difficult by adding objects in the occluded slots, and by providing the director with photographs that showed a mirror image of his perspective. This allowed the director to make occasional left-right "errors." We incorporated several such credibility cues, and, indeed, none of the participants suspected the confederate during the experiment. When given financial incentive after the experiment to correctly guess if their partner was a confederate or not, only 2 participants guessed post hoc that he was. Excluding the data of these 2 does not alter the pattern of results.

Materials

We used 12 different arrays, each of which included a perspective difference. In each array, there was an occluded object that the addressee might consider as being the intended referent in the critical instruction. For example, in the trial illustrated in Figure 1, the critical instruction referred to moving "the small candle." The intended referent was the left-most candle from the addressee's perspective, but the addressee also saw a smaller candle that was occluded from the director's perspective. In another array, there were three blocks: one on the top row, another on the second row from the bottom, and a third on the bottom row. The first two were visible to the director, but the one on the bottom was occluded. The critical instruction to "pick up the bottom block" referred to the block on the second row, but the addressee could have initially considered the occluded block as the referent. In the discussion that follows, we call the occluded object the occluded referent and the slot that contained it the occluded slot. From the addressee's egocentric perspective, the occluded referent was the intended referent.

Given that addressees' eyes tended to survey the objects, we included a control condition in which the object in the occluded slot was changed so that it was not a potential referent of the critical utterance (e.g., in the control condition corresponding to Fig. 1, a small toy monkey, rather than a candle, was occluded in the bottom row). Each participant received all 12 items, 6 in the experimental condition and 6 in the control condition, counterbalanced across participants. Items were presented in random order, with the constraint that no more than 2 items in the same condition follow consecutively. Items had two to four slots that were not visible to the director; the location of these slots varied between items.

Coding

We defined a temporal window of observation to code for eye fixations. The window started at the noun phrase that identified the shared referent (e.g., the expression "small candle"), and ended when the addressee identified the intended referent. We defined the point of final identification as the last fixation on the shared referent right before the addressee reached for that object. In the few cases in which the addressee did not look directly at the target object, the hand touch was considered the end of the window of observation. We counted a fixation on an object if the eye gaze remained in the slot for at least 100 ms consecutively. A coder who was blind to the hypothesis coded videotape for the end points of the window. In addition, a computer program used the digital information of eye fixation coordinates to determine the values of most of the dependent measures.

Results and Discussion

To see whether addressees considered referents that were inaccessible to the director, we first counted the number of fixations on the occluded slot throughout the observation window. On average, the addressees' eyes fixated the occluded slot almost twice as often when it contained a referent (test condition) than when it contained a non-referent (control condition), $t(20) = 2.7$, $p < .02$ (see Table 1). Similarly, the addressees' total fixation times within the occluded slot were 242 ms longer for the test condition compared with the control condition, $t(20) = 4.9$, $p < .001$. The data for eye gazes at the occluded objects, then, suggest that the directors were considering those objects as referents.

The temporal sequence of fixations was also indicative of the way addressees resolved the ambiguity. Figure 2 shows the time line of eye

Table 1. *Mean number of fixations on the occluded object and their mean summed duration*

Measure	Experiment 1		Experiment 2	
	Test	Control	Test	Control
Number of fixations	1.01 (1.16)	0.65 (0.91)	0.90 (0.98)	0.33 (0.65)
Total fixation time (ms)	420 (576)	178 (284)	452 (656)	106 (278)

Note. Standard deviations are in parentheses.
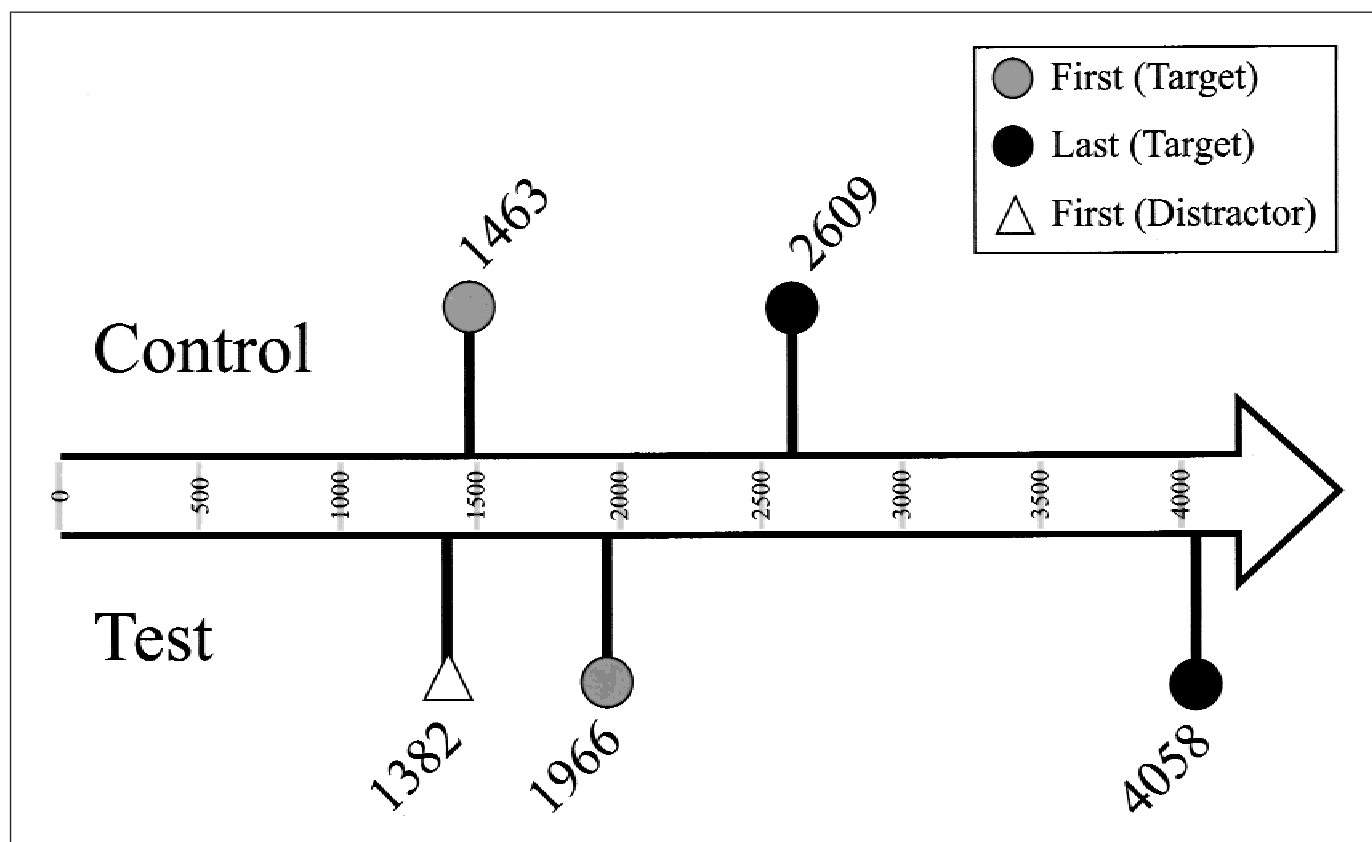
Fig. 2. Time line of eye fixations in Experiment 1, showing average latencies (in milliseconds) following the critical noun phrase (point 0).

fixations averaged over all items. The addressees' initial eye movements to the occluded object were the fastest, 584 ms faster than initial eye movements to the shared referent in the test condition. The initial fixation on the shared referent was delayed by 503 ms in the test condition compared with the control condition, $t(18) = 2.5$, $p < .05$, demonstrating that the egocentric interpretation interfered with the initial consideration of the shared object.

The final fixation on the target right before the addressee reached for it indicated the decision point. This fixation was delayed by 1,449 ms in the test condition compared with the control condition, $t(20) = 4.2$, $p < .0001$, suggesting interference in selecting the shared candle as the intended referent. Some of the final fixation latencies were quite long, so to make sure our results were not inflated by those data points, we also considered the median latency, which is not affected by outliers; it showed the same pattern: test, 2,767 ms; control, 2,067 ms; $t(20) = 2.9$, $p < .01$.

Though addressees noticed the intended object, as indicated by the initial fixation on it, they took longer in the test condition than in the control condition to decide that it was the intended referent, and to reach for it. This decision lag between the initial and final fixation was indeed enlarged from an average of 1,146 ms difference in the control condition to 2,092 ms in the test condition, $t(20) = 3.4$, $p < .01$.

The eye fixation data, then, demonstrate that addressees' knowledge about the perspective of the director played a role in error correction. Yet this correction was not always fast enough, and the addressee's hand actually reached for the occluded object in 23% of the trials in the test condition, compared with none in the control

condition, $t(20) = 4.4$, $p < .01$. In about a quarter of these cases (6%), the addressees recovered and corrected the erroneous hand reach, but in the rest of the cases (17%), the addressees grabbed the occluded object and moved it. Apparently, the egocentric interpretation can be so compelling that it has the potential to override the knowledge that the director cannot possibly see the occluded object.

The results of this experiment are surprising because they suggest that even when addressees clearly knew what information was shared and what was inaccessible to the director, they still occasionally used an egocentric interpretation strategy. But this interpretation is correct only if participants fully realized which objects were inaccessible to the director. Given the novelty of the methodology and the counter-intuitive nature of the results, we conducted a second experiment that attempted to ensure that addressees were fully aware that the objects were occluded.

EXPERIMENT 2

Forty native English speakers participated in this experiment. One participant guessed that the director was a confederate during the experiment, and 4 others guessed this only after the fact. The method, coding, and analysis were almost the same as in Experiment 1; we therefore describe only the differences between the two experiments.

Differences in the Method

Unlike in Experiment 1, the location of all occluded slots was identical for all items. Therefore, addressees did not have to identify

Fig. 3. Time line of eye fixations in Experiment 2, showing average latencies (in milliseconds) following the critical noun phrase (point 0).

the location of the occluded objects with each new trial. Also, we wanted to make sure that addressees knew which objects were not accessible to the director, so we instructed the addresses to help set up the arrays by hiding objects in the slots that were occluded from the director's perspective. With this procedure, it was absolutely clear to addressees which objects were shared and which were hidden.[2]

Experiment 2 used 10 of the 12 items from Experiment 1, dropping 2 items that might have been confusing. In addition, we replaced some of the noncritical objects and added 4 filler items. The confederate director in this experiment was female.

Results and Discussion

Overall, the results of this experiment were remarkably similar to those of Experiment 1. Addressees fixated the occluded slot almost three times more often in the test condition than in the control condition (see Table 1), $t(40) = 6.67$, $p < .001$. Fixations of the occluded slot were also 346 ms longer in the test condition than in the control condition, $t(40) = 6.14$, $p < .001$. These results suggest that addressees considered the occluded object as a potential referent in the test condition.

───────────

2. We also manipulated whether participants believed that the occluded objects were selected by the experimenter or by a random process. Given that this manipulation did not make a difference, we collapsed the data from these two conditions.

The time course of comprehension demonstrates that, on average, the occluded object was fixated 1,487 ms before the target object in the test condition (see Fig. 3). Also, the presence of an occluded referent delayed the first fixation on the target object by 1,045 ms, $t(40) = 3.92$, $p < .001$. In the presence of an occluded referent, the final fixation on the target was delayed by 1,783 ms, $t(40) = 4.95$, $p < .001$. Consequently, the decision lag between the first and last fixation was 738 ms longer in the test condition compared with the control condition, $t(40) = 2.71$, $p < .01$. These results are not determined by outlier latencies, as median latencies for the final fixation showed the same pattern: test, 3,300 ms; control, 2,600 ms; $t(40) = 4.73$, $p < .001$.

Much as in Experiment 1, in 20% of the cases addressees reached for (5%) or grabbed (15%) the occluded referent in the test condition, but they reached for the occluded object only once in the control condition. This high level of reaches toward the occluded objects is striking given that addressees hid those same objects in the occluded slots shortly before they rearranged the objects according to the director's instructions.

These data suggest that, indeed, addressees occasionally used an egocentric interpretation strategy. When this strategy led them to consider the wrong referent in the test condition, they had to use information about mutual knowledge, consider which objects the director knew or did not know about, and correct their interpretation. This correction delayed the final identification of the target referent.

We did find evidence that addressees attempted to focus on mutually shared objects before they heard the critical utterance. During a

window of 5 s before the critical utterance, they were more likely to fixate on shared than on occluded objects in the test condition (mean number of fixations = 0.18 and 0.05, respectively; $t[40] = 5.1$, $p < .001$), and the average fixation was longer on shared than occluded objects (means = 86 and 17 ms, respectively; $t[40] = 4.7$, $p < .001$). It appears that addressees were preparing to select the referent from among mutually shared objects. Yet despite this seemingly deliberative attempt, once they heard the director's next referring expression, addressees searched for a referent among both shared and occluded objects. This suggests that if mutual knowledge constrains comprehension, this constraint is partial at most.

GENERAL DISCUSSION

Both experiments demonstrate that addressees occasionally use an egocentric perspective when understanding referential expressions. That is, they tend to consider some referents from their own perspective, even when they know that these referents are inaccessible to the speaker. These findings have direct implications for assumptions about the role of mutual knowledge in comprehension and strategies for ambiguity resolution in general.

The Role of Mutual Knowledge in Comprehension

Our findings clearly demonstrate that addressees do not restrict their search for referents to mutually known objects. This is consistent with our earlier findings showing that when addressees' attention is distracted, they tend to disregard common ground in comprehension (Keysar, Barr, Balin, & Paek, 1998). But if comprehension is not restricted to common ground, what role does common ground play in comprehension? Our findings are consistent with two possible theories: one that assumes that mutual knowledge puts no constraint on comprehension and another that assumes a partial constraint on the initial phase of comprehension. Our experiments cannot distinguish between the no-constraint and the partial-constraint theories.

It is possible that mutual knowledge plays no role in the initial phase of comprehension. According to this possibility, addressees consider referents without regard to common ground. In this case, when they happen to identify a shared object as the referent, they would not need to consider mutual knowledge, but when they happen to select a nonshared object, they would need to use mutual knowledge to correct the error. According to this theory, mutual knowledge would be used by addressees only as a tool for error correction.

Another possibility is that mutual knowledge exerts a partial constraint on comprehension. According to this theory, addressees' knowledge that some objects are not shared reduces the probability that these objects would be considered as referents. Our evidence suggests that even though the probability of considering a nonshared object might be reduced, addressees often do consider these objects as potential referents. So, according to this theory, mutual knowledge plays two roles: It reduces the probability of considering a nonshared object, and it allows error correction when such referents are considered.

The Advantages and Disadvantages of an Egocentric Interpretation Strategy

When addressees clearly know which objects are shared and which are hidden, why do they still use an egocentric strategy? Such a strategy might require minimal cognitive effort because it uses accessible information and does not take into account alternative perspectives, which might involve further mental computation. Furthermore, this strategy takes advantage of the collaborative nature of conversation: Addressees can be lax in their use of common ground because errors can be detected and corrected by their conversational partners.

Yet errors are not always detected. There is a potential cost, then, to the egocentric interpretation strategy, as it occasionally leads one astray, sometimes irrevocably. Why would an addressee risk making such an error? One possibility is that the typical benefit of egocentric interpretation outweighs the typical cost of making an error. This explanation suggests that addressees "satisfice" when they comprehend (Simon, 1956, 1982; Tversky & Kahneman, 1974)—they use a strategy that is relatively effective though prone to errors, in order to accommodate a limited mental capacity.

Although the benefit of an egocentric interpretation might be the little demand it makes on the cognitive system, one needs to know more about the cost and the likelihood of making an error. Cost is difficult to estimate in the context of language comprehension, but there is reason to believe that the likelihood of an error due to egocentric interpretation is not very high. One reason for this assumption is that there is evidence that speakers tailor messages to addressees (e.g., Clark & Murphy, 1982; Fussell & Krauss, 1992; Horton & Keysar, 1996; Isaacs & Clark, 1987; Krauss & Fussell, 1991). Such "audience design" should reduce the likelihood that an egocentric interpretation would result in error. If this is true, then addressees might typically succeed in understanding the speaker's intention by relying on their own perspective.

CONCLUSION

Our findings are important in furthering the understanding of the mental processes that underlie conversation. Research in the field typically investigates comprehension from a macro-level approach and gives prominence to the assumption of shared perspective. Our micro-level investigation suggests that addressees tend to rely on information from their own perspective. Given that face-to-face interaction requires participants in conversation to act quickly, such a quick and easy comprehension strategy might be a crucial cognitive tool for resolving ambiguity in conversation.

REFERENCES

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

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

Clark, H.H., & Murphy, G.L. (1982). Audience design in meaning and reference. In J.F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.

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

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

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

Glucksberg, S., Krauss, R.M., & Higgins, E.T. (1975). The development of referential communication skills. In F.E. Horowitz (Ed.), *Review of child development research* (pp. 305–345). Chicago: University of Chicago Press.

Glucksberg, S., Kreuz, R.J., & Rho, S.H. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 323–335.

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

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

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

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.

Krauss, R.M., & Fussell, S.R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, *9*, 2–24.

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

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

Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Simon, H.A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

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

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