

PART I

Sociocultural Knowledge Influences



2

Making Sense of How We Make Sense: The Paradox of Egocentrism in Language Use

Dale J. Barr

University of California, Riverside

Boaz Keysar

University of Chicago

Language users routinely face the problem of making sense out of language: Speakers must design utterances that listeners can understand, and listeners must interpret utterances the way they were intended. Because ambiguity is pervasive in language use, pragmatic theories assume that speakers and listeners should strive to speak and understand against the background of a mutual perspective. However, our findings indicate that speakers and listeners are egocentric to a surprising degree. With respect to many current frameworks in psycholinguistics, these findings are anomalous: They suggest that language users are not properly designed for the task of making sense. Our goal in this chapter is to review these findings and to try to sketch out a new framework against which such egocentric behavior makes more (theoretical) sense. We propose that language users can rely on simpler mechanisms than current theories require because the work they are assumed to do to achieve a mutual perspective is actually distributed among processes in the language use environment.

The research presented in this book exemplifies how the meanings of figurative and other so-called “indirect” language can vary with social and cultural context. In this regard, it echoes a refrain that one encounters time and

again in the literature on language use—that meaning is underdetermined in the sense that the same string of words can convey anything from a benign comment to vicious sarcasm. One of the reasons why people continue to marvel at this idea is that we are still lacking a clear picture of how, in the face of vast ambiguity, people are able to make sense with and out of language. What we require is a better understanding of the mechanisms that underlie how people coordinate with each other.

One point on which all can agree is that speakers and listeners will only properly understand one another if they process utterances against a similar background set of assumptions and beliefs. In an influential study, Clark and Marshall (1981) characterized this background as *mutual knowledge*, the set of knowledge that is mutually held between interlocutors. Mutual knowledge is different from knowledge that is merely shared, because it is not only shared but is known to be shared. According to Clark and Marshall, mutual knowledge is essential because it is the only true guarantee of successful communication. They demonstrated this by showing that there are circumstances under which utterances that are based on information that two individuals have in common—but do not know they share—will fail. Because Clark and Marshall developed the theory extensively for the case of definite reference, it makes very clear predictions about how people should behave when producing and understanding referential expressions.

In definite reference, a speaker uses language to establish some object as the intersubjective focus of attention. Based on the cooperative principle, when speakers describe referents to listeners they should strive to be optimally informative (Grice, 1975). This means that they should provide no more and no less information than is necessary for the purpose of securing the listener's attention. Yet, what counts as "optimally informative" is defined by their mutual knowledge (Clark & Marshall, 1981). Imagine that Henry and Mabel are sitting at a table and there is a candle and a box placed between them. Henry wishes to get Mabel to put the candle in the box. What should he say? Clearly, he can simply tell her to "put the candle in the box." He should not say "the small candle" even if he is thinking of other larger candles that were on the shelf where he purchased the candle—these other candles are not part of their mutual knowledge. Similarly, Mabel will know that Henry is talking about the candle on the table and not the one that Ted bought for her yesterday, of which the small candle reminds her, because she knows that Henry does not know about Ted's candle. She will not ask "Which candle?" because she will know, by virtue of their mutual knowledge, that Henry is referring to the candle on the table.

As cognitive psychologists, we became interested in the mutual knowledge theory because of its strong implications for language processing. Even though much of the literature on mutual knowledge shuns explicit discussion of psychological mechanism, the theory's proponents have on

various occasions made clear assumptions that mutual knowledge should serve as a guiding principle in the access of information during language processing (Clark & Carlson, 1981; Gerrig, 1986) as well as in the organization of memory (Clark & Marshall, 1981) and the lexicon (Clark, 1998). Thus, speakers should design utterances with their specific addressee in mind (Clark & Murphy, 1982), and comprehenders should restrict the information they consider to mutual knowledge (Clark & Carlson, 1981). These proposals assume that language users can directly compute mutual knowledge on the fly and, therefore, call for the existence of powerful and efficient cognitive mechanisms that can guide the formulator and parser in their processing decisions. Building mutual knowledge directly into the language processing system seems advantageous not only because it would guarantee successful communication, but it would also reduce uncertainty and thus the complexity of the problem that language users face (Clark & Carlson, 1981).

From this vantage point, we were surprised to find from our own experiments on language use that speakers and listeners commonly violate their mutual knowledge when they produce and understand language. Thus, Henry will often refer to the candle as “the small candle” (Barr, 1999; Horton & Keysar, 1996), and Mabel will often consider the candle that Ted bought her as the referent or ask “Which candle?” even though there is only one candle that is uniquely defined by their mutual knowledge (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, Balin, & Paek, 1998). This behavior is egocentric because it is rooted in the speakers’ or listeners’ own knowledge instead of in mutual knowledge.

We have also observed effects in other domains that confirmed the generality of this egocentrism. Overhearers violate mutual knowledge in assessing whether an addressee will perceive an utterance as sarcastic (Keysar, 1994). Addressees interpret referential expressions according to naming precedents established by a previous speaker even though the current speaker was absent when the precedents were established (Barr, 1999; Barr & Keysar, 2002). Moreover, people turn out to be quite poor estimators of what others know. Speakers systematically underestimate the ambiguity and overestimate the effectiveness of their utterances (Keysar & Henly, 2002). When taught the meaning of an opaque idiom such as *the goose hangs high*, people overestimate the likelihood that others who are unfamiliar with the idiom will perceive its meaning (Keysar & Bly, 1995).

These findings might appear to conflict with other studies that purport to show effects of mutual knowledge on comprehension (e.g., Clark, Schreuder, & Buttrick, 1983; Greene, Gerrig, McKoon, & Ratcliff, 1994). However, Keysar (1997) noted that these studies suffer from a design flaw which confounds information that is mutual (i.e., known to be shared) with information that is simply known to the self (see also Lea, Mason, Albrecht,

Birch, & Myers, 1998). Once this flaw has been corrected, these methodologies yield the same systematic egocentrism as the studies just reviewed.

The evidence that language users are more egocentric than the mutual knowledge theory predicts is too abundant and too robust to be ignored. Yet, in light of the assumption that mutual knowledge is the only true guarantee of successful communication, it raises something of a paradox. How can language users be egocentric and still communicate effectively? In other words, how can people get away with being so alarmingly unsophisticated in dealing with the sophisticated problem of making sense?

We suggest that the only way to really make sense of this seemingly erratic behavior is to take a closer look at the circumstances under which it was designed to operate: the language use environment. Although certain aspects of the language use environment have been closely studied by researchers in disparate fields, there has been little effort to fit these pieces back together so that they can yield their full impact on theories of language use. In what follows, we examine the structure of the language use environment with the hope of showing how, counterintuitively, an egocentric way of speaking and understanding can make sense. Language users' egocentric behavior does not necessarily reflect a badly designed processing system; instead, we argue that it reflects the operation of simple heuristics that are adaptive given the normal circumstances of spoken language use. The burden of computing mutual knowledge is not one that individual language users carry alone; rather, it is one that they share by distributing it over processes in the environment. The richness of the environment enables language users to compensate for their limitations. In other words, language users have a bag of simple tricks that makes them look more sophisticated than they actually are. These tricks work because they are specifically designed to exploit the structure of the environment.

DOMAINS OF THE LANGUAGE USE ENVIRONMENT

What do we mean by the language use environment? We propose a typology that divides the environment into three subdomains: cognitive, interactional, and cultural (see Table 2.1). These domains are defined by the social and language units that constitute the primary units of analysis. By analogy, the set of domains can be construed as a set of lenses of varying power of magnification through which the theorist observes language use. Looking through the strongest lens, we can observe language use at its finest level of resolution, the cognitive domain. In this domain the basic social unit is the individual language user and the basic language unit, the single clause or utterance. This domain corresponds to the traditional level of inquiry in psycholinguistics, whose task is to understand the moment-by-moment processes underlying the production and comprehension of single utterances.

TABLE 2.1
Basic Domains of the Language Use Environment

<i>Domain</i>	<i>Social Unit</i>	<i>Language Unit</i>	<i>Characteristic Processes</i>	<i>Theoretical Import</i>
Cognitive	Individual	Clauses or utterances	Judgment and decision making Attention Memory Categorization	Places constraints on computation
Interactional	Dyad or group	Conversational turns	Epistemic exchanges Grounding (Clark & Brennan, 1991) Multimodal communication	Promotes shared understanding among the dyad or group
Cultural	Community	Languages	Establishment or diffusion of conventions	Promotes commonality of semantic representation among community members

The majority of our own research has primarily focused on understanding processes within this domain. What this research has shown is the existence of certain limitations on computation that prevent speakers and listeners from effectively deploying mutual knowledge when they process language. Any theory of how people establish shared understanding needs to take these limitations into account.

The broader domain in which the individual language user is embedded is the interactional domain, wherein the basic social unit of analysis is the dyad or larger group of conversational interactants and the basic language unit is the conversational turn. The focus at this level is on the interactive processes by which individuals manage the interaction (Sacks, Schelgloff, & Jefferson, 1974) and establish shared understanding (Clark & Brennan, 1991; Garrod & Anderson, 1987). Research on the interactional environment provides two extremely important insights. First, in conversation a shared perspective can be negotiated through an interactive process between interlocutors (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). Second, it shows that language as traditionally construed is really part of a multichannel system of communication (Clark, 1996) that includes paralinguistic information conveyed through the spoken channel

(Barr, 2003; Clark & Fox Tree, 2002) as well as body movements such as gestures that are conveyed visually (McNeill, 1992).

Much of what we currently know about the interactional domain is due to research by Clark and by other proponents of the mutual knowledge theory. In their view, interaction works at a metarepresentational level; that is, it is used as a vehicle for building models about what others know. Instead, we suggest that interaction works not indirectly at the metarepresentational but directly at the representational level in that it serves to coordinate individuals' conceptions of the discourse. This hypothesis is supported by some of our findings which indicate that people egocentrically apply what they have learned with one interlocutor when talking to the next one despite the fact that this violates mutual knowledge (Barr, 1999; Barr & Keysar, 2002). What is important about interactional processes, we argue, is that they greatly reduce the amount of work that must be done by processes in the cognitive domain.

Last, it is useful to remind oneself that communication is only made possible by the existence of shared semantic representations in the language community. The degree to which individual perspectives diverge or converge depends on the degree of overlap in how people in the community represent linguistic and world knowledge. With little overlap in semantic representations, communication will seem difficult. With large amounts of overlap, communication will seem effortless. To understand what is necessary for successful communication, it is important to evaluate the degree of overlap on which language users can typically rely. This requires an understanding the mechanisms that generate similarities or differences in semantic representation among the members of a language community.

Having completed this overview of the domains of the language use environment, we now put on our first set of lenses and take a closer look.

The Cognitive Domain

The reader might question our classification of the mind of the individual language user as a domain within the larger *environment* because this term has traditionally been used to refer to the set of information that is outside the mind of the individual. However, following Herbert Simon (1996) we view the structure of the human mind as a kind of an environment for thought. We propose that the cognitive domain is relevant to pragmatic inquiry because it imposes limits on the kinds of computations that language users can make during real-time conversation. Furthermore, the so-called "external" environment is itself made possible through the existence of cognitive structures that encode, store, and retrieve information. Therefore, accessing information while planning and interpreting utter-

ances will partake of such basic cognitive processes as judgment, categorization, attention, and memory, and it will thus reflect inherent features of how they operate.

Mechanisms of Decision Making in Language Use. Because utterances are inherently ambiguous, speakers and listeners face a high degree of uncertainty when they attempt to convey their own or decode another's intention on the basis of linguistic evidence. They may have insufficient information about what their interlocutor knows, or they may lack good evidence on which to base their assumptions. Our research shows that language use is no different from other domains in which people must make judgments under conditions of uncertainty. Specifically, we find that speakers and listeners employ the same anchoring and adjustment heuristic in understanding language as they do in other forms of problem solving.

When people make judgments to solve a problem, they tend to anchor their judgments in available information regardless of whether this information is actually useful for solving the problem. They then adjust away from this initial anchor, although the adjustment is typically insufficient; their ultimate response is skewed toward the initial anchor. For example, Tversky and Kahneman (1974) found that a group of high school students who were asked to estimate the multiplicative product of a sequence of numbers listed in descending order came up with higher estimates than a group who was asked to estimate the product of the same sequence with the numbers listed in ascending order. The explanation for this difference was that the students anchored their responses in the initial numbers of the sequence. Similarly, Epley and Gilovich (2001) showed that people often anchor on related, self-generated anchors and then make adjustments. When asked when Washington became president, they anchored on 1776, the lowest possible value, and then adjusted upward.

Likewise, research by Keysar and colleagues suggests that the same anchoring and adjustment principles apply to language processing, including how people perceive the meanings of figurative language (for a more extensive review of the anchoring and adjustment approach to language processing, see Keysar & Barr, 2002). Keysar (1994) showed that people's assessments of how others will perceive sarcasm are anchored in their own knowledge. In his study, participants read a passage involving two protagonists. For example, in one passage June recommends a certain restaurant to Mark. Mark goes there for dinner and has either a positive experience (i.e., enjoys the restaurant) or a negative experience (i.e., hates the restaurant). The next day he leaves a note for June that says, "The restaurant was marvelous, just marvelous." Participants were asked whether June would construe Mark's statement as sarcastic or sincere. Keysar found that respondents tended to think that June would perceive Mark's attitude even though she was missing the crucial information

about the valence of his attitude. The interpretation of this finding is that respondents anchored their estimate in their own understanding, and they failed to sufficiently adjust to the perspective of the uninformed addressee.

In addition, Keysar and Bly (1995) reported evidence that people anchor their assessments of how others will perceive the meanings of idioms in their own knowledge. In their study, they selected archaic English idioms such as *the goose hangs high* whose meanings were unfamiliar to modern day college students. The idioms were presented in the context of one of two passages by which the reader could infer either the original meaning or its opposite. For instance, in one passage, *the goose hangs high* was used to express optimism in the future. In a second passage, the same idiom was used to express foreboding about the future. Then, each participant read a passage in which a person used the idiom in conversation with a stranger but in a context that did not reveal its meaning. They were asked what they thought the stranger would take the idiom to mean. Sixty-two percent of respondents tended to think that the stranger would take the meaning to be the same thing they learned in the first place, whereas only 32% believed the stranger would understand the opposite meaning. In short, the respondents' anchoring of their judgments in their own knowledge of the meanings of the idioms led them to believe that the idioms were more transparent than they actually were.

In sum, these findings underscore the relevance of processes of cognition for theories of language use. Language users egocentrically anchor their judgments in available information and fail to fully adjust to the perspectives of others, just as they do in standard decision-making tasks. This suggests that the mechanism by which people assess shared perspective in speaking and understanding and the one they use in nonlinguistic problem solving are one and the same and, therefore, will be subject to the same sort of limitations.

The Control of Attention and Capacity Limitations. Attentional processes determine what information becomes accessible to cognitive systems in the normal course of their operation. We propose that to operate at a speed that is fast enough to cope with conversation, language processing systems are designed to operate on information that is made available by attention regardless of whether this information is part of mutual knowledge.

Many problems of pragmatics, such as formulating and disambiguating referential expressions, are problems that concern the control of attention. Speakers produce referential expressions to guide listeners' attention to referents. The expression that a speaker chooses in referring to some object—from an elaborate, full noun phrase to a simple pronoun—will depend on the degree to which the referent is in the focus of attention in the discourse (Ariel, 1998; Chafe, 1976; Gundel, Hedberg, & Zacharski, 1993). Likewise, listeners use the speaker's level of specificity as a cue to guide them in their

search for referents. For example, when listeners hear pronouns, they will assume that the referent is the current focus of attention.

However, under certain circumstances, a speaker's and a listener's focus of attention may not coincide. Suppose that your sister is telling you about her husband's trip to China, which her husband already told you about when you saw him yesterday. Feeling somewhat bored, your attention strays off, and you remember that you have not caught up with your friend Ben since he returned from Mexico. Just as you are thinking about Ben, your sister asks, "Have you talked to him yet?" The question is, will you consider Ben as the referent of the pronoun *him*, albeit temporarily, even though your brother-in-law has been established as the topic of discourse?

Because the attentional foci of speakers and listeners may not coincide, to avoid miscommunication it would seem optimal for comprehension systems to restrict the search for referents to mutually known information (Clark & Carlson, 1981). We put this restricted search hypothesis to the test by creating such a situation in our laboratory (Keysar et al., 1998). In this experiment, listeners wore a set of headphones while they helped a confederate director fill in missing details on a target picture (e.g., a picture of an airplane). At a critical moment, the listener's attention was distracted by a voice in the headphones that instructed him or her to look at a competitor picture. Just at the moment that the listener's eyes were focused on the competitor, the confederate asked a question about the target picture; for instance, "What color are its wings?" When the competitor picture that was at the focus of the listener's attention was a picture of a woman, listeners were faster to move their eyes back to the target picture than when it was a picture of a bird. This delay indicates that listeners mistook their own private thoughts as the referents of speakers' utterances. This is strong evidence against the restricted search hypothesis, and it supports our contention that language processing systems are designed to rapidly exploit the information made available by attentional processes regardless of its mutuality.

The finding that language processing systems do not initially restrict themselves to mutual knowledge suggests that misunderstanding will be systematic and pervasive. Yet, we know that people can compute the shared perspective when asked. Perhaps the reason why they fail to do this during routine language processing is because of capacity limitations; that is, the system must operate at such a time scale that it cannot accommodate inferences about mutual knowledge.

Horton and Keysar's (1996) study provides evidence in support of this view. In this study, participants described a target shape to a listener. The target shape was paired with a context shape that the speaker could see, but the listener could not. For instance, a target circle appeared next to a larger context circle that only the speaker could see. The question was whether the speaker would describe the target circle as "the small circle" or just "the

circle," with the former case representing a failure to consider the listener's lack of knowledge of the context circle. Horton and Keysar found that speakers were more likely to produce such egocentric utterances under the pressure of a response deadline than they were when they were allowed to respond at leisure.

We believe that this series of findings strongly implicates that language processing systems are designed to quickly settle matters of referential ambiguity by making rapid use of available information, regardless of its mutuality. Speakers and listeners routinely ignore even the most blatant cues to a referent's mutuality, such as whether it is occluded from the other's view, when they process utterances. In other words, language processing is anchored in the assumption that what is salient or accessible to oneself will also be accessible to one's interlocutor. Against the theoretical background in which mutual knowledge is taken as the only true guarantee of successful communication, the idea that the design of the language processor would embody such an assumption in its design is perplexing. Yet, we contend that although there are no guarantees of mutual understanding, there are powerful mechanisms in the interactional and cultural environments that promote shared perspectives among interlocutors, which can compensate for these limitations.

The Interactional Domain

One response that we sometimes encounter when we discuss our findings is that perhaps in real interaction, as opposed to in the psychological laboratory, people will be less egocentric. The assumption seems to be that to make interaction work, people will have to be more careful about adhering to mutual knowledge. Yet, we suggest that the opposite of this assumption may hold true: People are much more busy during real conversation than they are in the slow-moving and informationally rarified environment of the psychological laboratory; therefore, they will have fewer resources with which to constantly monitor what others know. However, because conversation is interactive and multimodal, it affords speakers and listeners copious feedback on their performance which affords them the opportunity to be more egocentric.

Even though conversational interaction is the fundamental setting of language use, the study of interaction has had little influence on the study of the psychology of language use. In many ways studies of conversational interaction developed in parallel with the laying of the foundations of sentence-level pragmatics in the Gricean and speech act traditions. The field of psycholinguistics, with its focus on single utterances and the isolated speaker, listener, or reader, has as its primary object the cognitive domain, and it has yet to gain the full impact of insights from studies of interaction.

One reason why research on interaction should cause people to rethink pragmatic theories is that it shows that pragmatic expectations are surprisingly flexible, such that speakers and listeners adapt them to suit their conversational experience. Consider the generalized pragmatic expectation that speakers should be optimally informative in designing referring expressions: They should provide their listeners with no more and no less information than is necessary (Grice, 1975). When choosing the word by which to refer to some object, speakers typically have a wide variety of choices available, each of which reflects different ways of categorizing the object (Brown, 1958). For instance, a shoe can be referred to as *the thing*, *the shoe*, or *the loafer*. Cruse (1977) suggested that speakers must determine the proper level of lexical specificity or their utterances can generate unintended conversational effects. For the purpose of definite reference, he argued that speakers should choose a specific term only when it is required, or they should default to a conventionally neutral level. For instance, if there are two shoes of different styles, the speaker might call the intended shoe *the loafer*. However, if there is only one shoe, the specific term *loafer* will sound especially “marked” and the speaker should default to the conventionally neutral, basic-level term *shoe*.

Brennan and Clark (1996) reported an experiment that challenges the generality of this pragmatic expectation. In their experiment, speakers played a game with listeners that required them to make reference to the same pictures over and over in contexts requiring different levels of lexical specificity. For instance, they referred to a shoe as *the loafer* multiple times because it appeared in the constraining context of another shoe. But when they were later presented with a test trial in which this same shoe appeared without the constraining context of the other shoe, they continued to use the specific term *loafer* even though it was now overly specific. Brennan and Clark took this as evidence that the dyad had established a mutually accepted “conceptual pact” to refer to this particular shoe as *the loafer*. Moreover, they found that speakers attempted to continue to use the overly specific term even when the test trial took place with a new listener who was absent when the pact was established.

We conducted an experiment that was similar to Brennan and Clark’s (Barr & Keysar, 2002, Experiment 3) except that we examined listeners’ expectations about the informativeness of speakers. Our experiment had a similar design, in which listeners learned to expect that a speaker would refer to a particular car as *the sports car* because it appeared recurrently within the constraining context of a station wagon. Likewise, the speaker established the precedent of referring to a particular flower as *the carnation* because it appeared in the context of a daisy. In a later posttest they saw a picture containing only the car and the flower and listened to a speaker refer to one of the objects. The question was, would listeners expect the terms

sports car and *carnation* which were overinformative in this context, or would they expect *car* and *flower* which were sufficient? What they actually heard was the word *car*. Note that if they expected to hear the overly specific terms, they would tend to mistake the word *car* as the initial phonemic segment of the word *carnation*. By tracking listeners' eye movements, we found that listeners expected the overly specific precedents, as revealed by more fixations to the carnation than in a baseline condition. More important, we found this expectation of the precedent to be equally strong among listeners who performed the posttest with a new speaker, one who could not have known of the precedent's existence.

The research on referring precedents is striking because it shows that interlocutors consider their conversational experience more important than prevailing norms of informativeness. Through interaction, speakers and listeners are able to adapt their language use to suit their own purposes. By relying on precedents, interactants reduce the complexity of the problem they face. Speakers will face fewer options during lexical selection; hence, listeners will benefit from less uncertainty in lexical identification. One reason why speakers might use naming precedents and listeners might expect speakers to use them is because they are part of their mutual knowledge. Another reason is because over the course of the interaction repeated use has simply made the lexical item strongly available and the underlying conceptualization strongly entrenched. These explanations differ in that the former assumes that interaction creates changes at the metarepresentational level, in terms of what is mutually accepted and mutually known, whereas the latter assumes that the changes are occurring directly at the representational level. The fact that speakers and listeners both use precedents in ways that violate mutual knowledge offer support for the latter interpretation. Furthermore, this interpretation is consistent with other research that shows how conversational interaction leads to conceptual convergence in how people represent the content of the discourse (Garrod & Anderson, 1987; Markman & Makin, 1998). To be clear, we do not dispute the idea that interacting speakers and listeners can and do keep track of what others know. But because interaction naturally causes speakers and listeners to similarly represent discourse information, we suggest that they need not always consult this metaknowledge to successfully communicate.

A second way in which interaction makes coordinating understanding easier is that it enables people to engage in what we call the *epistemic exchange*, an interactive exchange that serves as a proxy for the direct computation of mutual knowledge. We propose that certain interactive episodes in conversation represent instances of what Kirsh and Maglio (1994) termed *epistemic action*—cases in which people gain information about the world through direct action instead of through computing that information. In an epistemic exchange, speakers or listeners proceed on the basis of an egocen-

tric assumption and gather information about what their interlocutors know through feedback, even though they could have directly computed the partner's perspective. In the case of conversation, interaction provides interlocutors with ample opportunities to learn about what other people know or do not know without having to expend effort computing it themselves.

For example, imagine you are sitting across the table from Henry, and between you there are two candles that you both can see. In addition, there is a smaller candle that is obscured from Henry's view and is even smaller than the two mutually visible candles. When Henry tells you to "pick up the small candle," you could potentially compute the intended referent as the smaller of the two visible candles, because you know that he does not know about the even smaller hidden one. In Clark and Marshall's (1981) terms, only the visible candles are physically co-present and part of your mutual knowledge. However, you might simply pick up the smaller one without really thinking or feel confused and ask, "Which candle?" It might be easier (and more accurate) to get information about Henry's perspective from Henry himself, even though you could derive the identity of the referent on the basis of your mutual knowledge.

We revisited a set of data reported in Keysar et al. (2000) and looked for evidence that addressees engage in such epistemic exchanges. We found that under circumstances such as the one in this example, 27% of the time listeners performed the following actions: (a) asking for clarification (10%), (b) moving the small candle and then being corrected by the speaker (14%), or (c) both (3%). In contrast, in our control condition wherein the object corresponding to the hidden small candle was replaced with a nonreferent (i.e., a glass), in the vast majority of cases addressees were able to go right for the intended referent and required an interactive exchange only 6% of the time. In summary, even when addressees are presented with clear cues to what is mutually known, they often opt to resolve ambiguity by engaging in an epistemic exchange rather than computing the referent themselves. Especially given our finding that people are poor estimators of what others know, it makes a lot of sense for addressees to exploit the dynamics of interaction to distribute the burden of reference resolution rather than try to compute it themselves.

Another dimension of conversation that gives interaction its dynamism is multimodal communication. Despite the conventional terminology, a speaker does more than just speak and a listener does more than just listen. Speakers look and gesture as they speak. Listeners watch, nod, and make facial expressions as they listen. This background of multimodal activity provides interactants with a channel by which they can continually monitor their level of mutual understanding, and one that is backgrounded so as not to obtrude upon the official business of the conversation (Clark, 1996; Clark & Brennan, 1991). When speakers witness an uncomprehending look from

a listener midsentence, they can choose to elaborate on or repair their utterances. Listeners can nod and provide other back-channel information to show their continuing attention and that they are following the thread of the discourse. In other words, speakers and listeners not only communicate in the traditional sense but use the backgrounded multimodal channel to give constant evidence of their level of understanding. As with the interactive dimension of conversation, the multimodal character of conversation makes things much easier on language users.

Kelly, Barr, Church, and Lynch (1999) showed that listeners can read a speaker's pragmatic intention from their gestural behavior. In their study, viewers were presented with a video clip in which two actors acted out an everyday scenario that ended with a pragmatically ambiguous target utterance such as *it's hot in here*. The target utterance could potentially be construed as either a literal statement or an indirect request (i.e., to open the window). In one condition, speakers pointed to an object while delivering the target utterance, such as a closed window, that pertained to the intended meaning of the indirect request. In another condition, speakers kept their arms at their sides as they delivered it and maintained eye contact with the addressee. In the former condition, listeners were far more likely to interpret the utterance as an indirect request, showing that gesture can indexically ground the meaning of utterances. Yet, not only did the gesture disambiguate the speech, but the speech served to disambiguate the gesture—people were better at identifying what the speaker was pointing to when they heard the accompanying speech (which did not mention the object) than when they simply saw the pointing gesture. These findings suggest that speech and gesture work together to convey pragmatic meanings.

Another set of studies by Barr emphasizes how even a single vocal channel of communication can carry multiple dimensions of signals that enhance conceptual and linguistic coordination. In one study, listeners learned a set of novel color categories from a pretrained expert by viewing instances of each category and hearing prerecorded labels from the speaker (Barr, 2003). Listeners were able to detect the speaker's level of certainty in the classification from paralinguistic cues such as filled pauses (e.g., *um* and *uh*), hesitations, and rising or falling intonation. Given that people are more certain about the classification of typical than atypical instances, the speaker's paralinguistically conveyed certainty enabled listeners to differentiate good from bad examples and thereby facilitated learning of the categories.

Finally, these same kinds of paralinguistic cues can guide listeners in the identification of referents. Barr (2001) found that speakers produce different "hesitation signatures" when they formulate descriptions of new referents as compared to when they retrieve established precedents to refer to old referents. New referent signatures contained longer hesitations and were more likely to contain a filled pause. A follow-up comprehension experi-

ment found listeners to be sensitive to these signatures. When listeners heard a description of a new referent that was preceded by a hesitation signature containing a filled pause, they were over 300-ms faster at comprehending the description than when the filled pause was replaced by incidental noise. It is surprising that something as humble as an *um* can yield such a large advantage to comprehension, and such a finding suggests that listeners determine the identity of referents not only by what speakers say but by the apparent effort that they put into saying it.

In this discussion of the interactional domain of language use, we attempted to show how the interactive and multimodal processes of conversation simplify the coordination of mutual understanding by reducing the corresponding burden of processes in the cognitive domain. Although many of the same practitioners of the mutual knowledge theory have been the prime champions of research on interaction and multimodal communication, what differs is our interpretation of the implications of this domain for cognitive processing. Because these approaches have tended to eschew explicit discussion of mechanisms, they fail to fully appreciate the degree to which interaction serves as a proxy for the explicit computation of a shared perspective. Interaction serves as a vehicle not only for coordinating the metarepresentations of language users but for the representations themselves. It affords speakers and listeners the opportunity to engage in epistemic exchanges that help them learn about others' perspectives with a minimum of effort. Finally, the multimodal nature of conversation provides language users with immediate feedback and extra channels for communication. In short, the interactional domain enables language users to be more egocentric because it distributes the work that must be done to achieve mutual understanding over other processes in the environment.

The Cultural Domain

Whereas the interactional domain focuses on patterns of language use within a dyad or group of interactants, the cultural domain looks at language use through a wide-angle lens that encompasses the broader language community. We construe the culture as the repository of conventional practices that the members of a community have in common. The cultural domain is relevant to theories of pragmatics because it informs us about how much language users can take for granted when they interact with other members of their communities.

When interactants establish temporary patterns of language use to communicate effectively, they build on preestablished cultural patterns of language use—specifically, linguistic conventions—that the broader language community has shaped over time. The amount of work that language users will need to do to understand one another, whether through direct compu-

tation of mutual knowledge or through interactive exchanges, will depend on the degree to which they proceed from the same starting assumptions about the meanings of conventions. If everyone had exactly the same knowledge and experience, there would be little need for mutual knowledge because people could count on others knowing what they know. Conversely, if everyone had wildly different representations, communication would be extremely difficult. The need for communities to develop common representations among members presents a massive problem of social coordination. How do communities solve this problem? On what level of commonality of representation can two average members count? And how do these commonalities come about?

Traditionally, the field of pragmatics has sought to explain the establishment and use of linguistic conventions as a product of the accumulation of mutual knowledge among the members of a language community (Lewis, 1969). According to this view, individuals adhere to community-wide conventions because they have a preference to conform to the practices of other members of their community. Over the course of their experience, they build up a representation of how the modal member of their community will behave. Lewis stated, "If one has often encountered cases in which coordination was achieved in a certain problem by conforming to a certain regularity, and rarely or never encountered cases in which it was not, he is entitled to expect his neighbors to have had much the same experience" (p. 40). Mutual knowledge gives individuals a justification for conforming to the conventions in that it gives them reason to expect others to do the same. In essence, what this view assumes is that there can be no conventions without mutual knowledge.

Barr (in press) reported a series of multiagent computer simulations that shows that mutual knowledge is not necessary for the establishment and maintenance of semantic conventions in language communities. He argued that such representations of communal knowledge are unnecessary because conventions emerge as by-products of dyadic-level mechanisms of coordination. In the simulations, individual agents played a simple signaling game with other agents in the community. Each agent had a lexicon that mapped four symbolic forms onto four meanings. Initially, the form-to-meaning mappings were randomized for each agent. During each round of the simulation, each agent played the signaling game with a randomly selected partner from the community. The agent would attempt to communicate a random sequence of four meanings to its partner, and the partner would attempt to match the speaker's meanings. Agents received feedback as to whether they were correct or incorrect, but they did not receive any information about the other agent's mapping. Agents updated their mappings and then went on to play the next round with new partners.

Although these agents had no representation of what was going on in the community, but just adjusted their lexicons based on their experiences during a sequence of isolated interactions, the communities converged quite robustly to a single set of conventions. Even when they did not establish a single system, they typically converged on several spatially organized signaling systems or dialects, which are hallmarks of human language.

These simulations suggest that coordination can be achieved in language communities as a by-product of the work that language users do in the dyad rather than as the result of high-level calculations about what others know. In addition, it demonstrates that the work that individual language users do to coordinate with their language partners ultimately subserves the purpose of making their representations ever more similar to other members of their community. The fact that they can count on other members of their community having similar experiences with language greatly reduces the work they must do when they speak to others who are like themselves.

CONCLUSION

We began this chapter with a discussion of the paradox of egocentrism in language use. Although the underlying assumption of cooperation requires language users to speak and understand against the background of their mutual knowledge, language users appear to routinely disregard this knowledge. They expect an addressee to perceive sarcasm even when the addressee lacks crucial evidence about the speaker's attitude. They routinely consider hidden objects and private thoughts as the intended referents of speaker's expressions. They expect speakers to follow linguistic precedents that were established by another speaker. All of these findings show that language users are not designed to the exacting standards of pragmatic theories. According to these theories, language users' egocentric behavior does not make sense.

We argued that to make it make sense, we need to look not at the nature of the problem that language users must solve but at the structure of the environment in which language use is embedded. The environment can be divided into three subdomains, which is like looking at a single picture through lenses of different powers of magnification. Our research has focused on the cognitive domain, and it has uncovered limitations on language users' ability to effectively deploy mutual knowledge when processing single utterances. If we are to take these limitations seriously, then we must look elsewhere to find mechanisms that can compensate for these limitations.

The finding of egocentrism in language use need not imply that language users are somehow not adequately designed for the purpose of coordinating understanding. In fact, the processing system's rapid use of available information is exquisitely tuned to provide maximally efficient processing given

the interactional and cultural domains in which it operates. In essence, what we are claiming is that language users can get away with a large degree of egocentrism because the work of achieving shared understanding is distributed over interactive, multimodal, and cultural processes in the environment. Interactive processes such as the epistemic exchange allow language users to discover what other people know through negotiation without having to compute it themselves. Multimodal channels of communication enable language users to simultaneously convey and perceive multiple dimensions of meaning, including online feedback about their level of understanding. Cultural processes serve to indirectly coordinate the representations of the individuals in a community, greatly reducing the work they must do in the dyad. In essence, language users can be simpler than pragmatic theories require because part of the burden of coordination is distributed over the domains of dyadic and communal interaction.

To be clear, we do not intend to cast doubt on the possibility of mutual knowledge either as a theoretical construct or as a factor that is operative in conversation. Yet, we wish to emphasize that whether or not a speaker or listener uses mutual knowledge on a particular occasion is an empirical question. The coincidence of perspectives is not a fortuitous event but rather a direct consequence of the operation of background processes in the language use environment. Thus, the mere observation that a speaker produces an utterance that is in alignment with mutual knowledge does not warrant the inference that she or he directly computed that knowledge as mutual at any time. The speaker may have or may have simply used information that was simultaneously available and salient to him or her and the interlocutor.

Moreover, our purpose is not to question people's ability to compute metarepresentations about what others know. Yet, our results imply that the cognitive system is designed to operate efficiently on representations, not metarepresentations. However, this does not preclude the use of such higher order representations to monitor and correct problems that arise during conversation. Thus, speakers' and listeners' ability to compute mutual knowledge might be operative primarily when communication fails. According to our findings, it appears that mutual knowledge is most likely to be implemented as a mechanism for detecting and correcting errors instead of an intrinsic, routine process of the language processor.

Faced with the findings reviewed so far, we think it is important to rethink exactly what it means to be cooperative, a concept that is at the heart of most theories of language use. For one, the supposition that speakers strive to be maximally informative in lexical selection does not seem to fit what they actually do. Perhaps a better description of what they do is simply rely on their past and current discourse experience and select the term that is most strongly available to them. As the simulations by Barr (in press) show, one's own conversational experience can often be a reliable guide to what is

conventional in the community. That said, there are probably social politeness norms that will cause speakers to be more careful about their lexical selection (e.g., whether to refer to a woman as *my mother*, *mom*, or *Mrs. Smith*). Just when and how speakers will heed social circumstances during lexical selection are questions that warrant further investigation.

The presence of epistemic exchanges, hitherto unnoted in the literature, also calls for us to rethink the notion of cooperation. If speakers and listeners really had as strong expectations of cooperation as the theory says they should, then they would routinely be confronted with behavior that seemed uncooperative. Perhaps language users tolerate some slack over the short run because it is the most effective way to share the burden of coordinating understanding over the long run. Perhaps it is not through the individual sentence by which language users demonstrate they are cooperative, but rather it is how they behave over the course of the conversation.

In closing, the inherent ambiguity of language creates a complex problem for language users and seems to call for correspondingly complex mechanisms that enable language users to successfully communicate in the face of such vast uncertainty. Our message is that this complexity should not be sought within the head of the individual language user but rather in the cognitive, interactional, and cultural domains that comprise the environment of language use. Language users can be simpler than theories require because the environment is more complex than these theories envision. It is only by considering the behavior of language users against this background that we can truly make sense of how we make sense.

REFERENCES

- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24, 65–87.
- Barr, D. J. (1999). *A theory of dynamic coordination for conversational interaction*. Unpublished doctoral dissertation.
- Barr, D. J. (2001). *Paralinguistic correlates of discourse structure*. Paper presented at the 43rd annual meeting of the Psychonomic Society, Orlando, FL.
- Barr, D. J. (in press). The emergence of conventions in language communities. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. Cambridge, MA: MIT Press.
- Barr, D. J. (2003). Paralinguistic correlates of conceptual structure. *Psychonomic Bulletin & Review*, 10, 462–467.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1482–1493.
- Brown, R. (1958). How shall a thing be called? *Psychological Review*, 65, 14–21.

- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 26–55). New York: Academic Press.
- Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Clark, H. H. (1998). Communal lexicons. In K. Malmkjær & J. Williams (Eds.), *Context in language learning and language understanding* (pp. 63–87). Cambridge, England: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–49). Washington, DC: American Psychological Association.
- Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 313–330). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73–111.
- 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, 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.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, 1–39.
- Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics*, 13, 153–164.
- Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Divergent processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12, 391–396.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Gerrig, R. J. (1986). Process models and pragmatics. In N. E. Sharkey (Ed.), *Advances in cognitive science* (Vol. 1, pp. 23–42). New York: Wiley.
- Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, 33, 511–526.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.
- Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59(1), 91–117.
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4), 577–592.

- Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, 26(2), 165–208.
- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24, 253–270.
- 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, England: Cambridge University Press.
- 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., & Bly, B. (1995). Intuitions of the transparency of idioms: Can one keep a secret by spilling the beans? *Journal of Memory and Language*, 34, 89–109.
- Keysar, B., & Henly, A. S. (2002). Speakers' overestimation of their effectiveness. *Psychological Science*, 13, 207–212.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, 18, 513–549.
- Lea, R. B., Mason, R. A., Albrecht, J. E., Birch, S. L., & Myers, J. L. (1998). Who knows what about whom: What role does common ground play in accessing distant information? *Journal of Memory and Language*, 39, 70–84.
- Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, 127, 331–354.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, 696–735.
- Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.

