

Less Is More: A Minimalist Account of Joint Action in Communication

Hadas Shintel,^a Boaz Keysar^b

^a*Gonda Multidisciplinary Brain Research Center, Bar Ilan University*

^b*Department of Psychology, The University of Chicago*

Received 11 June 2008; received in revised form 19 December 2008; accepted 5 February 2009

Abstract

Language use can be viewed as a form of joint activity that requires the coordination of meaning between individuals. Because the linguistic signal is notoriously ambiguous, interlocutors need to draw upon additional sources of information to resolve ambiguity and achieve shared understanding. One way individuals can achieve coordination is by using inferences about the interlocutor's intentions and mental states to adapt their behavior. However, such an inferential process can be demanding in terms of both time and cognitive resources. Here, we suggest that interaction provides interlocutors with many cues that can support coordination of meaning, even when they are neither produced intentionally for that purpose nor interpreted as signaling speakers' intention. In many circumstances, interlocutors can take advantage of these cues to adapt their behavior in ways that promote coordination, bypassing the need to resort to deliberative inferential processes.

Keywords: Referential communication; Common ground; Audience design; Memory associations; Interlocutor-specific effects

1. Introduction

Language is inherently social. Though it can serve as a medium for cognitive processes within the individual mind, language is fundamentally a social practice that rests on shared understanding and use. Similar to basketball and other activities that require coordination of the behaviors of individuals, language use can be viewed as a

Correspondence should be sent to Hadas Shintel, Gonda Multidisciplinary Brain Research Center, Bar Ilan University, Ramat-Gan, Israel. E-mail: hshintel@gmail.com

form of joint activity (Clark, 1996) that involves the coordination of meaning between individuals.

How is coordination of meaning achieved? One possibility is that coordination is driven primarily by deliberative processes that require interlocutors to consult their theory of mind and infer their conversational partner's mental states—her beliefs, desires, and intentions. This model of the partner is then called upon in planning one's own utterances and in interpreting the partner's utterances. Specifically, interlocutors consider their common ground, that is, information they believe to be mutual and available to both (see Clark & Marshall, 1981), and use it to constrain production and comprehension. Speakers presumably engage in "audience design": They design their utterances taking into account the common ground with specific addressees. Listeners, in turn, interpret utterances while assuming that speakers engaged in audience design (Clark & Murphy, 1982).

It is easy to see how partner-specific adaptations can provide a powerful method for minimizing miscommunication. However, such elaborate reasoning that requires interlocutors to keep updated metarepresentations of the other's beliefs that are separate from their own representations of the situation is both time consuming and cognitively demanding (Epley, Keysar, VanBoven, & Gilovich, 2004; Horton & Keysar, 1996; Rossnagel, 2000).

In contrast, minimalist explanations of joint action do not assume that successful coordination entails consulting a model of the interlocutor. Rather, coordination can emerge out of domain-general processes taking advantage of the multiplicity of cues available in interaction. Unlike signals intentionally produced as communicative signs and reflecting partner-oriented adaptations, such cues need neither be produced intentionally to adapt to addressees' informational needs nor interpreted with respect to their potential function as signaling speakers' intention.

In this paper, we will propose that general cognitive processes that are not unique to communication, are nonstrategic in the sense that they are not performed intentionally for communicative functions, and do not involve the attribution of mental states can give rise to meaning coordination in communication. We will focus on two kinds of evidence. First, we will discuss adaptations that appear to be tailored to specific interlocutors, given their individual or social identity, and review evidence suggesting that such adaptations may be a consequence of nonstrategic memory processes. Second, we will discuss various cues that can benefit comprehension without being specific to interlocutors' informational needs. We suggest that such nonstrategic processes can support the coordination of meaning and lead to communicative patterns that appear as if they are sensitive to interlocutors' intentions and needs. Our aim in this paper is not to provide a comprehensive review of the role of common ground and audience design in language processing (see, e.g., Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000; Nadig & Sedivy, 2002; see Barr & Keysar, 2007 for review). Rather our point here is that in many circumstances, coordination of meaning can proceed through simpler means, bypassing mental state inferences.

2. Adapting to a particular interlocutor

2.1. Speaker-specific effects in comprehension

Objects can be referred to in various ways; the same object can be called “the shoe,” “the loafer,” “the suede shoe,” or even “Fido’s toy.” With repeated reference, interlocutors come to ‘entrain’ on specific expressions and use them subsequently (Garrod & Anderson, 1987). Having referred to a shoe as “the loafer,” interlocutors will continue calling it “the loafer,” even when other expressions may identify it more readily given the context. Indeed, speakers continued using the same expressions even when these became overinformative for their addressees (Brennan & Clark, 1996). Listeners, in turn, expected speakers to continue using the same overinformative expressions. For example, after a speaker called a flower “carnation” to distinguish it from a daisy, listeners expected her to continue using “carnation,” rather than the optimally informative basic-level term “flower,” even when there was no other flower (Barr & Keysar, 2002).

Indeed, established expression-referent mappings can reduce referential ambiguity and consequently facilitate comprehension. Violation of such precedents, on the other hand, incurs a processing cost. Interestingly, the cost associated with precedent violation appears to be speaker specific. Metzing and Brennan (2003) found that listeners were delayed in identifying referents when speakers violated their own precedents and used new referring terms. In contrast, when new speakers used these terms, listeners showed no delay. Moreover, Keysar, Barr, and Lim (2001) found that the cost associated with speakers violating their precedents was not mitigated by the fact that the new expression had previously been used by another speaker and thus was already mapped to a specific referent. These findings indicate that the cost associated with precedent violation does not result merely from the effort required to establish a new mapping.

Such speaker-specific adaptation could promote coordination of meaning, because it makes speakers’ behavior more predictable. Various factors can affect how speakers refer to things, including the availability of lexical forms, the way speakers think about those things, and the pragmatic context. For instance, *shoe* is more available than *loafer*, it is a different way of conceptualizing a shoe than *Fido’s toy*, but it cannot distinguish a particular shoe from other shoes. In principle, speakers could be constantly changing the way they refer to things as a function of varying circumstances. Instead, speakers reuse referring expressions even when they are not contextually optimal (Brennan & Clark, 1996), allowing listeners to rely on precedents to constrain interpretation and predict what speakers will say.

While it is clear that listeners rely on precedents, it is not clear why. One possibility is that listeners use precedents to infer speakers’ communicative intentions. If Kate first used “the loafer” and then made reference to “the suede shoe,” listeners may reason that with this change she intended to signal to them a change in referent, rather than simply violated a precedent she established with them (Metzing & Brennan, 2003). If change in terminology is taken as a signal, then it must be considered a signal only if the addressees are expected to know the terminology that the speaker originally used. On this account, the cost associated with an actual violation of precedents is not only speaker specific, but it is specific to

interactions within a speaker–listener dyad, or at least to interactions in which precedents are common ground. If Kate had called a shoe “the loafer” when talking to John, he will expect her to reuse the same expression when interacting with him, but not with other people. Thus, the cost of precedent violation should be listener specific as well as speaker specific.

In contrast, a minimalist account does not assume that listeners rely on precedents because they have a communicative signal value. Listeners expect speakers to be consistent, independent of who they are talking with. When speakers are inconsistent in their use and violate their precedents, listeners may incur processing costs irrespective of any consideration of speakers’ communicative intentions. If John knows that Kate had called a shoe “the loafer,” he will expect her to continue calling it “the loafer” even when she is talking to Sarah, who never met Kate before. On this account, this expectation is directed at a specific speaker, but it is not constrained to a speaker–listener dyad.

These two accounts therefore make different predictions regarding the circumstances in which a processing cost will be observed. According to the former account, the processing cost is *partner* specific. It is specific to a speaker–listener dyad, and it is a function of the speaker’s history of use in interactions in which both speaker and listener acted as ratified participants. In contrast, according to the minimalist account, the processing cost is *speaker* specific. It is specific to a speaker given her history of use but is independent of the particular speaker–listener dyad. Because most studies manipulated the speaker’s identity while keeping the listener constant, speaker specificity and partner specificity could not be unfounded.

To distinguish the two, we conducted an experiment in which we manipulated the addressee’s identity and kept constant participants’ knowledge of speakers’ precedents (Shintel & Keysar, 2007). Most important, we manipulated participants’ belief regarding the mutuality of the precedents independently of their availability to participants to avoid a common confounding of mutuality of information with its availability (Keysar, 1997). Participants completed a referential communication task in which they followed the instructions of a confederate female speaker, while their eye gaze was being tracked. We prepared videos of the speaker performing the task with a confederate addressee while repeatedly using the same expression to refer to objects. For example, the speaker called an elephant-shaped baby rattle “the elephant rattle.” We manipulated mutuality of precedents by manipulating participants’ beliefs regarding the speaker’s knowledge. In the first phase, some participants interacted with the actual speaker, thereby establishing precedents as common ground. Other participants watched the prepared videos alone and were told the speaker does not know that they had watched the videos, thereby establishing that precedents were not common ground. Next, all participants interacted with the same speaker in person. On critical trials, the speaker referred to target objects either by using the precedent (e.g., “the elephant rattle”) or by using a new expression, thereby violating the precedent (e.g., “the baby rattle”).

If the cost of precedent violation is partner specific, addressees should rely only on common ground precedents. Only then could the listener infer that a change in expression is intended to signal a change in referent, because only then would listeners believe that the

speaker expects them to notice *the change*. In contrast, if listeners' knowledge of the speaker's history is sufficient to produce an expectation that she will adhere to precedents, there should be a processing cost when the speaker switches from "elephant rattle" to "baby rattle" independent of listeners' beliefs about the speaker's knowledge.

The results showed that listeners were delayed in identifying referents when they encountered new expressions; when listeners heard "the baby rattle" they tended to look at other provisionally potential referents, for example, a baby shampoo. This delay occurred whether precedents were mutual or not, supporting the idea that speaker-specific, rather than partner-specific, expectations underlie the use of precedents.

A control condition, in which the speaker used the same new expression but no precedent had been established, confirmed that the delay can be attributed to precedent violation rather than to the need to establish a new mapping. Precedent violation resulted in a reliable delay compared to when the same expressions were used in the absence of a precedent, even when precedents were not common ground. So listeners do not take change in terminology as a signal meant for them. Knowledge of the speaker's history of usage is sufficient to create expectations regarding her future linguistic behavior. A deliberative process of inferring the intended referent based on the speaker's communicative intentions is not only unsupported by the evidence, it is in fact unnecessary to bring about such expectations.

2.2. *Interlocutor-specific memory associations*

Listeners expect speakers to be consistent in their use, even when the speakers' history of usage is not common ground. How is this expectation of consistency realized? Such expectation need not be conscious or explicit; in fact, in postexperimental questioning only one participant reported noticing that the speaker used different expressions, suggesting that the effect was not mediated by consciously noticing the change. One potential explanation is that speaker-specific effects emerge out of memory representations that incorporate episodic information, such as the identity of the speaker who established the expression-referent mapping (Metzing & Brennan, 2003). Various models of memory maintain that memory depends on the degree of overlap between memory traces and memory probes and that contextual associations that are encoded in episodic traces can subsequently cue memory (e.g., Hintzman's MINERVA 2, Gillund & Shiffrin's SAM model for recognition, Pike's Matrix model, see Raaijmakers & Shiffrin [1992] for review). The speaker can function as one such retrieval cue, resulting in greater accessibility of an established expression-referent mapping when the speaker is the same as in the original context of encoding. The activated mapping can interfere with the establishment of a new mapping between the same referent and a new expression. In this sense, the speaker acts in a qualitatively similar fashion to other retrieval cues.¹ The effect of the speaker's identity may result not from a consideration of the speaker's communicative intentions, but from the salience and significance of information about the speaker during encoding, allowing the speaker's identity to act as a rather potent retrieval cue.

Support for the idea that memory associations can promote coordination comes from evidence of the role of listener-specific associations in language production. Horton (2007)

had participants produce category exemplars in response to category cues (e.g., “a fruit”) given by two confederates and then complete a picture-naming task in the presence of each confederate. Naming of objects that appeared in the exemplar task was modulated by the identity of the listener who was present during the picture task: Participants were faster to name objects previously associated with the present confederate. Furthermore, facilitation of naming was observed for members of categories associated with the present listener, even when these differed from the specific exemplars introduced in the exemplar task. Thus, the mere presence of an individual primed information associated with that individual, even in the absence of an intention to communicate. In fact, as the naming task did not involve any interaction, confederates did not act as conversational *partners* in that task.

In Horton’s (2007) study, the presence of the other person primed information associated with that person without conscious recollection of that association, suggesting that people were not trying to retrieve information associated with that person. These results are consistent with the finding that amnesics who communicated about tangram figures were able to retain the final concise labels they had used (Duff, Hengst, Tranel, & Cohen, 2006). This was in stark contrast to their severe explicit declarative memory impairment and to the fact that they did not seem to keep an explicit record of the interaction with the communicative partner. Together, these results suggest that an explicit record of the expressions used by one’s interlocutor, or even an explicit record of one’s own usage, is not necessary for coordination of reference.

Interestingly, in an eye-tracking comprehension study, Barr and Kronmüller (2004) found that implicit, but not explicit, sources of speaker information constrained reference resolution. The experimenter initially introduced the objects known to either a male or female speaker (e.g., “you have a buckle, right?”). When the speakers later referred to those objects, listeners were more likely to initially look at a competitor object that was phonologically related to the target (e.g., a *bucket*) than at an unrelated object, even when it was not common ground and known only to the other speaker. On the other hand, if objects were initially introduced by the speaker who knew about them (e.g., “I have a buckle”), and therefore were encoded in that speaker’s voice, then looks to a phonological competitor that was not common ground did not differ from looks to an unrelated object. Importantly, in both conditions participants knew which objects are known to each speaker. However, explicit information about common ground did not constrain initial processing in the absence of effective speaker–object memory associations that could provide the foundation for speaker-specific effects.

Because the strength of association in memory between a speaker and speaker-relevant information does not necessarily correlate with the degree to which this information is common ground, we would expect to also find the converse pattern: If listeners have strong memory associations between a speaker and speaker-relevant information, we may observe speaker-specific adaptations reflecting these associations even when information is not common ground. This is indeed what we found in our study (Shintel & Keysar, 2007). Given the relatively strong association between a speaker and expression-referent mappings established after watching the speaker repeatedly use the same expressions, knowledge that these mappings are not common ground did not reduce their effect on comprehension.

Our claim here should be distinguished from the claim that memory demands modulate the degree to which production and comprehension reflect partner-specific adaptations, such that adaptations are less likely when keeping track of common ground is especially demanding (Horton & Gerrig, 2005; S. Lin, B. Keysar, & N. Epley, unpublished data). Here we argue that because such adaptations are supported by general memory cues, not by consideration of the other’s communicative intentions, they are not constrained to conversational partners.

One might still argue that in the Shintel and Keysar (2007) study participants’ memory did not allow them to distinguish expressions that were not common ground (i.e., introduced in the video phase) from those that were common ground (i.e., introduced in the test phase). Horton and Gerrig (2005) argued that interlocutors’ behavior can reflect audience design only when appropriate memory representations are available. Obviously, if listeners cannot remember when precedents were introduced, their behavior may not reflect the common ground even if they did consider it. In this case, participants should show evidence of using common ground when the associated memory load is reduced. Thus, the cost of precedent violation should gradually increase over the course of the test phase as more expressions are introduced, increasing memory confusability. However, reanalysis of our data does not support this interpretation. An ANOVA with mutuality of precedents manipulated between subjects, and expression (old, new-violation, new-no precedent) and Order (first half vs. second half) manipulated within subjects, revealed only a significant effect of expression, $F(2,70) = 23.26$, $MSE = 352344$, $p < .0001$. No other effects were significant (all F s < 1). The cost of violating a precedent that was not common ground did not increase over time (see Fig. 1). In fact, a reliable cost was evident already in the first round of the test phase even though at that point there should have been little room for confusion with the preceding video phase.

Overall, these findings support the idea that listener- or speaker-specific adaptations of behavior are mediated by general memory processes rather than by communication-specific

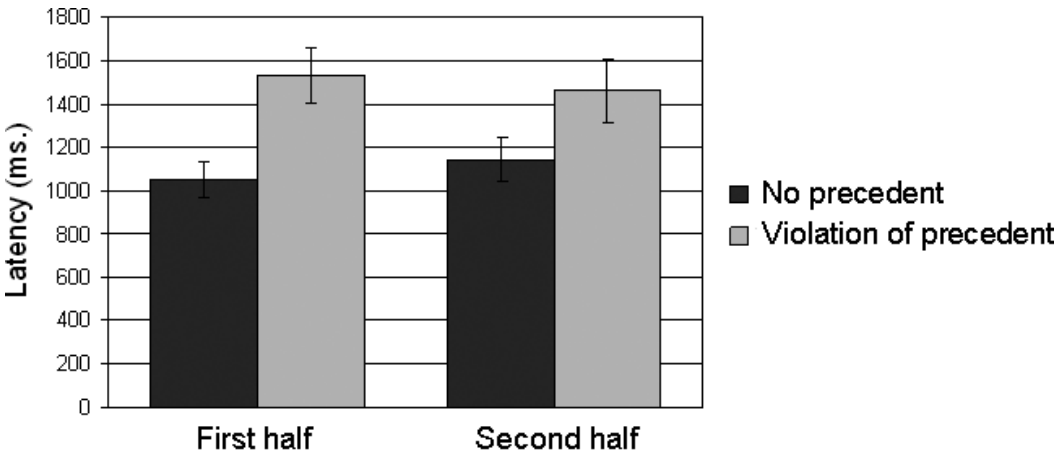


Fig. 1. First fixation latencies for noncommon ground precedents.

strategic adaptations (Horton & Gerrig, 2005²). Although some situations may give rise to genuine partner-specific adaptations, the findings reviewed here suggest that often effects that appear *as if* they are partner specific are realized by processes that are not sensitive to common ground.

2.3. *Effects of the interlocutor's social identity*

So far we have focused on memory representations associated with particular individuals and on their role in promoting interlocutor-specific adjustments of linguistic-communicative behavior. However, a great deal of our knowledge does not concern particular individuals but rather groups of individuals, based on shared ethnicity, nationality, hobbies, etc. Information about community membership can guide inferences regarding others' knowledge and provide a valuable source of information for interlocutors to draw upon (Fussell & Krauss, 1992; Isaacs & Clark, 1987). One can use "Citicorp building" to refer to a New York landmark when talking to New Yorkers, but probably not with non-New Yorkers. Obviously, the use of a proper name will make sense only if addressees know the name; in contrast, when addressees are unfamiliar with a place's name, a description is necessary. Although reference provides an obvious example, adaptations of communicative behavior to individuals by virtue of their group membership can occur at other levels as well. For example, listeners perceived vowels differently if they believed the speaker was from Detroit than from Canada, based on the speech stereotypes associated with that identity (Niedzielski, 1999).

While the processes underlying the effects of perceived group membership on communicative behavior are still not clear, they may be partly mediated by memory associations based on social identity. Categorical information about salient social attributes such as race or gender, has been shown to affect memory source attributions (Taylor, Fiske, Etcoff, & Ruderman, 1978): In a "who said what" paradigm, individuals made more intra-category errors (within racial or gender groups) than inter-category errors. In the previous section we discussed the potential role of memory associations based on individual identity in promoting adaptations of communicative behavior. Similarly, memory associations encoding social identity may play a role in promoting adaptations of communicative behavior based on interlocutors' community membership.

Support for this idea comes from recent studies in our lab (Senay & Keysar, in press) that examined the effect of gender on how listeners interpreted the violation of precedents. Participants first heard two speakers describe nonsense figures. For example, one speaker described a figure as "the sitting man," while the other described it as "the flying ghost." Critically, the two speakers were either of the same gender or of a different gender.³ Next, participants played a referential communication task in which they selected the figure described by the female speaker, the director, who either used the same description or "borrowed" the other speaker's description. If the director called a figure "the flying ghost" after she originally called it "the sitting man," participants were more likely to select a different figure. So participants expected the speaker to adhere to her precedents and interpreted new terms as implying different referents. However, this was contingent on speakers'

gender. If the speaker borrowed the term from another female, participants were more likely to select the original figure, as if they treated precedents set by one female as applicable to other females. The same pattern emerged when participants recalled the descriptions used by each speaker: Participants were more likely to misattribute to a speaker the descriptions of a same-gender speaker. Finally, when gender was made more salient, the likelihood of same-gender misattribution increased, suggesting misattribution was modulated by gender, not merely by the similarity between individuals irrespective of their gender. These results suggest that precedents are associated in memory with gender categorical information.

More broadly, these results suggest that adaptations of communicative behavior based on social identity can emerge out of memory associations that encode social categorical information. The results reviewed here focused on the consequences of such representations for source misattribution. However, they can also lead to behavior that is more in sync with addressees' perceived communicative needs, even when it does not reflect adaptations explicitly tailored to these needs.

3. Nonstrategic generic-listener adaptations

We have suggested that communication can benefit from nonstrategic adaptations of behavior to specific interlocutors. These need to be distinguished from generic-interlocutor adaptations that do not vary with the characteristics of the interlocutor or the communicative situation but can promote comprehension in general (Dell & Brown, 1991). It is this category of adaptations which we now turn to consider.

In general, studies can fail to find support for audience design in two ways: (a) when speakers do not provide disambiguating information even though their addressees need it (e.g., Arnold, Wasow, Asudeh, & Alrenga, 2004), or (b) when speakers provide disambiguating information that their addressees do not need (e.g., Dell & Brown, 1991; Kraljic & Brennan, 2005). While the first category of findings poses a problem for communication, the second category shows that various useful cues may be available to listeners, even when such cues are not intentionally produced to facilitate communication. Such cues can be used for disambiguating the linguistic signal, bypassing the need to engage in resource-demanding inferences about the interlocutor's mental state and intentions.

Face-to-face interaction involves visual and acoustic information that provides powerful cues for intention, cues that are not necessarily intended as part of communication. Eye gaze, for instance, serves as a powerful cue to attention (Langton, Watt, & Bruce, 2000). There is evidence of a tight coupling between eye movements and processes underlying language production and comprehension. For example, speakers typically gaze at objects while preparing their names (Griffin & Bock, 2000), and listeners typically gaze at referents of utterances (Cooper, 1974; Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). Eye gaze can even trigger reflective shifts of attention when it is task irrelevant (Langton et al., 2000). So eye gaze can indicate what speakers and addressees have in mind, even when it is not meant or interpreted as an intentional communicative signal. Ostensive-communicative cues such as eye gaze may have a special role in interaction. For example, infants interpret

actions differently depending on the presence of ostensive cues (Gergely, Egyed, & Király, 2007). However, this tendency reflects a general bias in processing information conveyed by eye gaze rather than an inference involving the attribution of mental states (*ibid.*, see also Csibra, 2003). So eye gaze may provide an external cue to intention that circumvents reflective, more cognitively demanding processes of explicitly considering the other's mental states. In this way, eye gaze information may constrain the referential domain and support coordination while conserving cognitive resources.

Eye gaze can even be a more powerful cue than explicit knowledge. Hanna and Brennan (2007) found that addressees used the speaker's gaze in identifying the target before the linguistic point of disambiguation. When speakers' and addressees' displays were incongruent, addressees initially oriented in the direction of the speakers' gaze but later drew on their knowledge of the incongruency to adjust their gaze toward the target. This suggests that when interlocutors do share visual perspective, they can rely on a faster and simpler process that does not require them to explicitly use eye gaze as a basis for inferring the speaker's intention.

Indeed, Bard, Anderson, Chen, Nicholson, Havard, and Dalzel-Job (2007) found that when speakers needed to intentionally seek out information about listeners' gaze, gaze was not used consistently. Participants guided "followers" along a map route and received feedback about followers' gaze. Though they could have used followers' gaze as an indication of a mistake, they did not attend to followers' gaze to off-route landmarks. Furthermore, participants tended to rely on the less diagnostic verbal feedback rather than gaze feedback, a tendency that was exacerbated under time pressure. When the use of gaze information entails constantly monitoring for, and keeping an updated model of, addressees' understanding, speakers may rely instead on listeners' feedback and attempt repair when needed (Pickering & Garrod, 2004).

Other modalities also provide cues that can support the coordination of meaning. Speakers express meaning not only in words, but also in the gestures accompanying speech and listeners can glean information conveyed uniquely in gestures (McNeill, 1992; see Goldin-Meadow & Wagner, 2005 for review). Although gesture can communicate information, speakers also gesture to simply fulfill their own cognitive needs (e.g., Cook, Mitchell, & Goldin-Meadow, 2008; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). Speakers gesture even when they know that their gestures cannot be seen, as when they gesture when they talk to blind listeners or when a screen blocks the listener's view (Alibali, Heath, & Myers, 2001; Iverson & Goldin-Meadow, 1998; but see Alibali et al., 2001). Similarly, semantic information can be conveyed through a mapping of visual information onto acoustic properties of speech. For instance, when speakers described the direction of a dot's motion they modulated their speech rate and pitch height in a way that corresponded to the dot's speed and vertical direction (Shintel, Nusbaum, & Okrent, 2006). They did this even when the modulation was task irrelevant, the task involved no interaction, and speakers did not know that utterances would be presented to listeners.

Finally, cues may become informative without inferences about mental states of the speaker. Consider speech disfluencies such as *Sandra bumped into the busboy and the uh uh waiter told her to be careful*. When disfluencies appear at the beginning of clause

boundaries, as in this example, they facilitate parsing (Bailey & Ferreira, 2003). This could indicate that listeners use the disfluency to infer the speaker's underlying mental state. However, external noise that occurred at the clause onset had the same effect, suggesting the effect is driven by statistical learning. Statistical learning is a powerful general mechanism that allows listeners to capitalize on regularities in the input to constrain interpretation, without resorting to inferences regarding underlying mental states.

4. When is coordination difficult?

Language is inherently ambiguous, making it necessary to disambiguate utterances to decode the speaker's intention. Yet under many circumstances, general cognitive processes that do not involve a model of the interlocutor can support meaning coordination while conserving the resources necessary for mental states inferences.

Models of coordination in conversation have suggested different ways in which the cognitive burdens of dialog are distributed across participants (see Bard et al., 2007; for review of the different positions). However, cognitive load can also be offloaded by being tuned to the external constraints imposed on production and comprehension by the communicative context. To the extent that interlocutors share a context, they can rely on it to constrain ambiguity without resorting to resource-demanding, deliberative inferential processes. Visual context offers a clear example: Coordination is easier when interlocutors can indeed rely on shared visual context (Clark & Krych, 2004). However, without the support of shared context, individuals will need to resort to more complex processes, making coordination more difficult even when they are aware of the discrepancy in their perspectives (Keysar et al., 2000).

Even when interlocutors' representations initially diverge, they may converge via processes that do not invoke a model of the interlocutor. Pickering and Garrod (2004) suggested that interlocutors' representations become aligned because of the parity of representations used in production and comprehension and the priming between speakers and listeners. Though priming alone may not be sufficient for coordination, the interactive alignment account suggests a mechanism that leads to convergence of representations without requiring intermediary decision stages or a model of the interlocutor.

We have mainly focused on self-prompted adaptations, generated by individuals regardless of their interlocutor's feedback (see Barr & Keysar, 2007). But of course, interlocutors can also rely on feedback to detect temporary ambiguity or misunderstanding. Although individuals can attempt repair without consulting a model of the interlocutor (see Pickering & Garrod, 2004), it is also possible that once miscommunication becomes evident, individuals will be more likely to consider their interlocutor's informational needs in attempting to recover from miscommunication. Furthermore, explicit feedback may facilitate inferences about interlocutors' knowledge and thus reduce the computational demands involved in using a model of the interlocutor.

Finally, individuals may adapt to addressees' informational needs when they can rely on simple perceptual cues and when these cues are linked to clear behavioral adaptations.

A. Galati and S. E. Brennan (unpublished data) suggested that speakers can rely on a “one-bit” model that is based on a single cue and marks addressees’ needs as belonging to one of two alternatives. For example, speakers can use a single cue (new/old listener) and express new information more clearly. Such adaptations do not require people to keep an elaborate updated model of their interlocutor or consult this model on an utterance-by-utterance basis.

Reliance on these processes will alleviate the need to engage in more complex inferential processes. But it may be at a price of miscommunication, temporary or long-lasting, when context is not shared. Individuals may resort to more effortful strategies that involve taking into account the perspective of their interlocutor in specific circumstances, for example, when the cost of miscommunication is higher, or when cultural norms reduce the effort by providing people with routine experience with taking the other’s perspective (Wu & Keysar, 2007). Barring this, people seem to coordinate meaning with little reflection on the other’s mental states by relying on powerful, simple, and cheap cues. In many situations, we coordinate meaning seamlessly precisely because we are minimalists.

Notes

1. Of course, information about a partner may draw more attention and consequently be remembered better.
2. Horton and Gerrig (2005) suggest that individuals are trying to assess common ground, but are constrained by memory (e.g., p. 141). In that sense, their account differs from ours.
3. A production study indicated that descriptions were equally likely to be produced by both genders.

Acknowledgments

We thank Susan Brennan and two anonymous reviewers for their helpful comments and suggestions on a previous version of this paper.

References

- Alibali, M. W., Heath, D., & Myers, H. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, 51, 55–70.
- Bailey, K. G. D., & Ferreira, F. (2003). Disfluencies affect the parsing of garden-path sentences. *Journal of Memory and Language*, 49, 183–200.

- Bard, E. G., Anderson, A. H., Chen, Y., Nicholson, H. B. M., Havard, C., & Dalziel-Job, S. (2007). Let's you do that: Sharing the cognitive burdens of dialogue. *Journal of Memory and Language*, 57, 616–641.
- Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, 46, 391–418.
- Barr, D. J., & Keysar, B. (2007). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 901–938). New York: Academic Press.
- Barr, D. J., & Kronmüller, E. (2004). *Can talker identity constrain spoken word recognition?* Poster presented at the 45th annual meeting of the Psychonomic Society, Minneapolis, MI.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(6), 1482–1493.
- Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- 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. Joshi, 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. LeNy & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.
- Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesture makes learning last. *Cognition*, 106, 1047–1058.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 813–839.
- Csibra, G. (2003). Teleological and referential understanding of action in infancy. *Philosophical Transactions of the Royal Society of London B*, 358, 447–458.
- Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the “model of the listener.” In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics* (pp. 105–129). Hillsdale, NJ: Erlbaum.
- Duff, M. C., Hengst, J., Tranel, D., & Cohen, N. J. (2006). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, 9, 140–146.
- Eberhard, K. M., Spivey-Knowlton, M., Sedivy, J., & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409–436.
- Epley, N., Keysar, B., VanBoven, 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. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62, 378–391.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27, 181–218.
- Gergely, G., Egyed, K., & Király, I. (2007). On pedagogy. *Developmental Science*, 10, 139–146.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516–522.
- Goldin-Meadow, S., & Wagner, S. M. (2005). How our hands help us learn. *Trends in Cognitive Science*, 9, 234–241.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57, 596–615.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49, 43–61.

- Horton, W. S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, 22, 1114–1139.
- Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, 96, 127–142.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26–37.
- Iverson, J. M., & Goldin-Meadow, S. (1998). Why people gesture as they speak. *Nature*, 396, 228.
- Keysar, B. (1997). Unconfounding common ground. *Discourse Processes*, 24, 253–270.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–38.
- Keysar, B., Barr, D. J., & Lim, S. (2001). *Does conversational grounding involve partner-specific pacts? An example of the adaptive use of language*. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.
- Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology*, 50, 194–231.
- Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, 4, 50–59.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory & Language*, 49, 201–213.
- Nadig, A. S., & Sedivy, J. S. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, 13, 329–336.
- Niedzielski, N. A. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18, 62–85.
- Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral & Brain Sciences*, 27, 169–225.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology*, 43, 205–234.
- Rosnagel, C. (2000). Cognitive load and perspective taking: Applying the automatic-controlled distinction to verbal communication. *European Journal of Social Psychology*, 30, 429–445.
- Senay, I., & Keysar, B., (in press). Keeping track of speaker's perspective: The role of social identity. *Discourse Processes*.
- Shintel, H., & Keysar, B. (2007). You said it before and you'll say it again: Expectations of consistency in communication. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 33, 357–369.
- Shintel, H., Nusbaum, H. C., & Okrent, A. (2006). Analog acoustic expression in speech. *Journal of Memory and Language*, 55, 167–177.
- Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, 36, 778–793.
- Wu, S., & Keysar, B. (2007). Cultural effects on perspective taking. *Psychological Science*, 18, 600–606.