Language, Perspective, and Memory

RACHEL A. RYSKIN, SI ON YOON, and SARAH BROWN-SCHMIDT

Abstract

The ability to take the perspective of another person is ubiquitous in many everyday cognitive activities. In particular, it allows people to communicate efficiently with conversational partners. Speakers tailor what they say based on the listener’s knowledge and, likewise, listeners use what they know about the speaker to better understand what the speaker means. In this essay, we review foundational research on the role of perspective-taking in the domain of language processing and describe new lines of work that are beginning to explore the memory processes that support the efficient use of perspectives in conversation. We then discuss key avenues for future research, such as investigating whether the type of perspective-taking involved in creating memory reminders draws on the same underlying cognitive processes as in the domain of language processing. Exploring this interface between language, perspective-taking, and memory will require interdisciplinary crosstalk and integration of methodologies across the domains of memory and language research.

INTRODUCTION

Perspective-taking is the ability to mentally represent the beliefs and knowledge of another person, which may or may not differ from one’s own. More generally, the capacity to appreciate alternative perspectives is a key contributor to a variety of human cognitive activities—from being able to read a map (Figure 1) to interacting with another person. Consider that when any two individuals come together, they have a certain amount of knowledge about the world that is shared between them. For two Americans sitting in a ball-park watching a game, this joint knowledge or common ground would include popular cultural references such as the name of the President of the United States, as well as immediately available information in the context such as information about the last play, or the score in the game. These individuals would also have a certain amount of information that is private to each individual, and not shared with the other person, for example, what each one of them ate for breakfast. A central goal of many interactions is sharing some of this private knowledge, sometimes called privileged ground, in
order to introduce new information and keep the conversation interesting
(after all, a conversation is bound to be painfully boring if it only involved
repeating to one another what is already jointly known).

Keeping track of what others do and do not know is an ability that is
observed even in very young children. For example, when young children
see an actor place a toy inside a container and leave the room, they expect
the actor, on her return, to search for the toy in that same container even
if the child knows that the toy was removed from this container. Thus, the
child is able to distinguish their privileged knowledge that the toy is no
longer in the container, from the actor’s false belief that the toy is in the
container. This ability to distinguish self-knowledge from other-knowledge
allows the child to generate the prediction that the actor will search for her
toy in the original location. The ability to represent the fact that the actor
has a false belief about the location of the toy—a belief that differs from the
child’s own—is present by the second year of life, if not earlier (Baillargeon,
Scott, & He, 2010).

Perspective-taking is likewise a key contributor to adult interpersonal com-
munication; effective speakers tailor how they speak to their audience, and
effective listeners adjust their understanding of what is said based on what
they know about their interlocutor (Clark, 1996). For example, when describ-
ing New York City landmarks, New Yorkers will use proper names (e.g.,
Rockefeller Center) when speaking to other New Yorkers and descriptive
phrases (e.g., the building with the flags in front of it) when speaking to a person who has never been to New York (Isaacs & Clark, 1987). Similarly, an effective listener might interpret a word such as *tweet* with different meanings, depending on their beliefs about what the speaker knows. If a college student were listening to a young person, *tweet* might bring to mind the social media site, “Twitter”; if they were talking to an elderly person, the same word might bring to mind a bird.

This ability to adjust our expectations about the actions of others and to tailor how we use language with other individuals critically depends on representations of the perspective of other individuals. In order to appreciate that another person has a different perspective than one’s own, for example, individuals must store in memory a representation of their interlocutor’s knowledge state that is distinct from their own knowledge state. They must also be able to access those memory representations quickly, when cued appropriately. The nature of these memory structures and the ways in which they are accessed during language production and comprehension are areas of active research. The products of this research have implications for a variety of cognitive domains.

**FOUNDATIONAL RESEARCH**

Representations of the perspective of others play a central role in face-to-face communicative settings. In conversation, the common ground between the interlocutors is thought to be the fundamental backdrop against which communication takes place (Clark, 1996). As a result, many of the advances in our understanding of how perspective is represented and used come from psycholinguistic studies of how language is processed in face-to-face conversation. This research shows that both speakers and listeners alike tailor language processes based on the knowledge that is jointly held with their conversational partner.

**COMMON GROUND IN LANGUAGE PRODUCTION**

It is well known that when we refer to something in the world, we must distinguish what we intend to describe from other potential referents (Olson, 1970; Osgood, 1971). When one orders a doughnut at a favorite doughnut shop, it simply will not suffice to say “One doughnut, please.” Instead, the speaker must specify her referential expression “one doughnut” to pick out the one she wants, from many that she does not want, as in “One glazed chocolate cake doughnut, please.” Thus, speakers design their referring expressions with respect to the physical and cognitive environment they find themselves in. Consistent with the logic that common ground is central to
language use (Clark, 1996), speakers also change their utterances depending on their beliefs about what their conversational partner knows and does not know. For example, imagine a situation in which a speaker will ask a listener to hand her one of several items in a display, such as the smaller of two triangles in Figure 2. In a situation where both the speaker and listener see the two triangles, speakers typically specify the size of the triangle, as in “Please hand me the small triangle” (Figure 2a). The adjective “small” plays a critical role in this linguistic exchange, because it specifies which of the two triangles the speaker is referring to. However, if the larger triangle is not visible to their partner (Figure 2b), speakers often will refer to it simply as “the triangle” because the contrasting adjective, “small,” becomes superfluous (Nadig & Sedivy, 2002; Wardlow Lane, Groisman, & Ferreira, 2006) and potentially confusing. From the listener’s perspective, there is only one triangle. Thus, speakers use their knowledge about the listener’s perspective in the situation, in order to constrain what they say.

While the shared physical environment is one component of common ground, another mechanism by which common ground is formed is through communication itself. A central finding in studies of conversational language is that, as we converse, the ways in which we refer to various topics of discussion become more definite and succinct over the course of a conversation (Wilkes-Gibbs & Clark, 1992). For instance, when describing an ambiguous black-and-white image such as the leftmost object in Figure 3, on a first attempt, a speaker might say, “it resembles someone who looks like they’re trying to climb stairs. There’s two feet, one is way above the other.” When redescribing the same figure later in the conversation, the speaker might now refer to it more concretely and efficiently, as in “the stair climber.” Memory for the previous mention of that ambiguous figure, and the knowledge that this memory is shared with one’s conversational partner, allows the speaker to refine her phrasing. This phenomenon is observable even on a global level. Consider that in the early days of the Internet, commercials enticed potential customers to view their product on the “world wide web,” whereas nowadays, we simply refer to the “web.” This shortening of the expression

<table>
<thead>
<tr>
<th>Listener</th>
<th>Listener</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="a" alt="Diagram" /> Listener</td>
<td><img src="b" alt="Diagram" /> Listener</td>
</tr>
</tbody>
</table>

**Figure 2** (a) Both triangles are visible to the listener. (b) A barrier blocks the larger triangle, making it hidden from the listener’s perspective.
represents the growth in global common knowledge about the Internet, and, as such, indicates that, as a global community, we have formed common ground for this concept.

Key evidence in support of the argument that these phenomena represent common ground between people (and not simply private knowledge, e.g., about what the web is) comes from findings that speakers no longer use these shortened terms such as stair climber or web when speaking with a person who is unknowledgeable about the term. In such circumstances, speakers revert to longer, more descriptive phrases (Horton & Gerrig, 2002; Wilkes-Gibbs & Clark, 1992). This ability to tailor referential expressions to the knowledge of specific addressees is remarkably flexible and powerful. Speakers are able to keep track of a large number of distinct and intermixed items, some of which are shared with one partner, and some of which are shared with a different partner (Horton & Gerrig, 2005). Speakers are also able to switch back and forth between conversational partners with whom they share distinct knowledge, appropriately adjusting their expressions to be more descriptive when the current addressee is “naïve” with respect to that given item, and using established naming conventions when the current addressee is familiar with the convention (Horton & Gerrig, 2002; Yoon & Brown-Schmidt, 2014).

COMMON GROUND IN LANGUAGE COMPREHENSION

In much the same way as speakers tailor how they speak depending on the knowledge of their addressee, listeners, too, engage in complementary, partner-specific adjustment during conversation. Once two people in a conversation have converged on a particular expression (e.g., “the shiny cylinder”) to refer to a given referent, listeners are surprised to hear the same speaker switch to an equally plausible but novel expression (e.g., “the silver pipe”), suggesting that the novel expression provides a less effective cue to the representation of that referent stored in memory (Metzing & Brennan, 2003). However, if a new person joins the conversation, these novel expressions are no longer surprising, and are processed more quickly.

Figure 3  Example tangrams.
The fact that the penalty for novel expressions does \textit{not} extend to speakers who do not share common ground for the original term, shows that this is not simply an egocentric process. Instead, listeners take into account what knowledge their partner does and does not have when generating an interpretation of what they say.

Listeners similarly use what they know about a speaker’s \textit{visual} perspective within the physical context when interpreting language (Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002). For instance, if a listener is looking at a display that includes two red triangles, only one of which is visible to the speaker (Figure 4), when the listener hears an instruction such as, \textit{“Click on the red triangle,”} the listener will look primarily at the red triangle in common ground. The listener is able to rule out the triangle in privileged ground because she knows that, from the speaker’s perspective, only one triangle is visible. Complementary findings include evidence that if a speaker asks you to hand her something, e.g., \textit{“Hand me the cake mix …,”} listeners interpret the speaker as requesting the cake mix that is out of her own reach, and \textit{not} a cake mix that she could have reached from her position; after all, she would not need to ask for it if it were within reach (Hanna & Tanenhaus, 2004). Further, when interpreting informational questions, for example, \textit{“What’s above the …,”} listeners readily interpret them as asking about something that the speaker does not already know—otherwise, there is no need to ask a question.

\textbf{Figure 4}  
Listener’s view while listening to the instruction, “Click on the red triangle.” The gray background indicates an object that is visible to the listener, but not to the speaker.
in the first place (Brown-Schmidt, 2012; Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). Taken together, this body of research shows that listeners access representations of the speaker’s perspective, and incorporate this information into the unfolding interpretation of the speaker’s utterances.

One way of conceptualizing this sensitivity to partner-specific common ground is as a type of contextual integration, in which information, such as a jointly established description of a referent (e.g., *the shiny cylinder*), is bound to the person with whom the information was experienced. Framed in this way, understanding how this binding of partner identity with information from the conversation is encoded, maintained, and accessed in the service of language processing becomes a central question. Research in the memory tradition distinguishes concepts such as item memory, including memory for a word on a list, from the source of that information, including the room the words were studied in (see Johnson, Hashtroudi, & Lindsay, 1993, for a review). While *source* in the memory literature is conceptualized broadly as the conditions in which information is encoded, in the domain of conversational language, perhaps the most important source is the identity and perspective of the person with whom you are conversing. Consistent with this idea that interlocutors are particularly good contextual cues comes from a meta-analysis of environmental context effects in memory research (e.g., studies in which the locations of study and test are either the same or different), which reported a large effect of whether the *Experimenter* stayed the same from study to test (Smith & Vela, 2001). In what follows, we discuss new research that furthers our understanding of the role of context in a discourse and explores the neural underpinnings of discourse representations. We then describe potential domains of inquiry that could be pursued to fruitfully expand our understanding of the link between memory and language as it pertains to perspective-taking.

**CUTTING-EDGE RESEARCH**

The literature on language processing in conversation has clearly established that information in the local context guides language use, including visually available information (Olson, 1970), as well as information about what a partner does and does not see or know (Hanna *et al.*, 2003; Nadig & Sedivy, 2002). While it is clear that conversational partners encode and use information about content that is discussed, such as the way particular referents are described (Wilkes-Gibbs & Clark, 1992), less well established is the extent to which contextual information, including information about source, is encoded and maintained in conversation. Indeed, source memory is often more fragile than item memory (e.g., Ferguson, Hashtroudi, & Johnson, 1992).
An important question then is whether memory for source can be cued effectively at later time points in a conversation, such that future referring is altered as a result of previous contextual information, or whether the change of temporal and cognitive context is enough to render the source information inaccessible. Moreover, what are the memory systems that support the encoding and maintenance of item and source information to be accessed in the discourse?

**Past Contexts and Future Language Use**

According to the context-memory view of discourse (Yoon & Brown-Schmidt, 2013), speakers maintain memory for how they had described previous referents in the conversation. These memory representations guide not only how speakers re-refer to these referents (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992) but also how they describe other, previously unnamed referents in the future (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013). Key evidence for persistent effects of past contexts on future referring comes from experiments that examine how the way in which speakers refer to a given referent in one context, shapes how the same item is referred to in a new context. For example, if a pair of conversational partners are presented with a display that contains several different fish pictures (Figure 5a), and are asked to refer to each picture, partners will establish brief but descriptive labels for each fish (e.g., “the curved round fish”) that distinguish among the category exemplars in the scene (i.e., the other two fish).

Evidence of persistent context effects comes from findings that even when the context changes and now includes only a single category exemplar (Figure 5b), speakers continue to use previously established expressions (e.g., “the curved round fish”) even though they are overly specific given the immediate context. In other words, while the abbreviated expression the fish

![Figure 5](https://example.com/figure5.png)

**Figure 5** (a) Speakers say “the curved round fish” to disambiguate the target (circled) fish from the other two fish. (b) Speakers say “the curved round fish” to disambiguate the target fish from the contrasting fish seen previously (lexical entrainment). (c) Speakers say “the glittery fish” to disambiguate the novel target fish from previously discussed fish (lexical differentiation effect).
would have uniquely identified the referent, speakers persist in using the longer, more descriptive term that had been established previously (Brennan & Clark, 1996). According to the context-memory view, the source of this effect is that the speaker’s representation of the relevant context includes both the single fish in the immediate visual context as well as multiple other contrasting fish from past contexts.

Distinguishing this context-memory view from a simpler account on which speakers had simply encoded the term the curved round fish as the name of the referent (rather than encoding both the term and the items in the past context), is supported by evidence that previous referents shape the descriptions of future referents. One demonstration of this is the phenomenon of lexical differentiation (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013), in which speakers differentiate new discourse referents from similar referents discussed in past contexts. For example, imagine a different situation in which conversational partners first refer to the fish in Figure 5b as the fish. In a subsequent context in which a different, novel fish is present (Figure 5c), speakers are more likely to describe it with a modified noun phrase (e.g., glittery fish) compared to a situation where they had not previously described other fish (Van der Wege, 2009). This lexical differentiation effect shows that speakers consider how distinct referents from the discourse history were previously described when designing new expressions. One intriguing finding is that while Yoon and Brown-Schmidt (2013) observed lexical differentiation in language production, there was no evidence that listeners expected speakers to differentiate current from past referents. Instead, listeners interpreted established terms such as fish quickly, regardless of whether the term had previously been used to describe a different fish or not. This suggests that the scope of the relevant discourse context for speakers and listeners may differ, resulting in subtle asymmetries in the relationship between reference and context.

**Memory Systems in Language Use**

The persistence of the historical discourse context in language use raises the question of how this information is encoded in memory. At least some evidence suggests that multiple memory systems may be involved. In particular, neuropsychological work with individuals who have severe declarative memory impairment (amnesia) due to bilateral hippocampal damage shows patterns of deficits and sparing of function consistent with contributions of multiple memory systems (Duff & Brown-Schmidt, 2012). The idea that the hippocampus might be involved in processing partner-specific representations of common ground falls naturally from other findings pointing to hippocampal involvement in the processing of object–location relations, even
over short time-scales (Hannula, Tranel, & Cohen, 2006; Hannula & Ranganath, 2008). For example, individuals with amnesia are largely successful at using information about what a conversational partner can and cannot see in a visual display to constrain interpretation of referring expressions (Rubin, Brown-Schmidt, Duff, Tranel, & Cohen, 2011). However, when the same individuals were presented with situations in which common ground was established linguistically (and no visual cues to what was shared and what was not shared were present), they were no longer able to distinguish common ground from privileged ground after a brief delay. Converging evidence from studies of the ability to process short narratives reveals significant deficits in the ability of individuals with amnesia to track discourse referents over the course of a story, and use information about their relative salience to process pronouns that refer back to these characters (Kurczek, Brown-Schmidt, & Duff, 2013). Yet, other evidence shows sparing of learning. Specifically, individuals with amnesia successfully develop talker-specific representations of the accents of different talkers, and use these representations to guide speech perception (Trude, Duff, & Brown-Schmidt, 2014). Together, these findings emphasize the need to understand the memory systems that contribute to and are necessary for the ability to tailor language use to one’s conversational partner.

KEY ISSUES FOR FUTURE RESEARCH

The literature on language processing in conversation has clearly established that better understanding the relationship between language and memory processes is a key area in need of development, not only in terms of experimental findings but also in terms of theory building and the design of implemented models of the underlying processes. In what follows, we describe several new and promising lines of inquiry that address these questions.

According to one recent proposal, the cognitive processes involved in the design of a referring expression for another individual are analogous to those in play when one is constructing a cue that will be used later to access a specific event in memory (Tullis, 2013; Tullis & Benjamin, 2014). In a cue-generation task, learners are given a large set of target words and told to generate one or more cue words for each one. These cues are then presented to the participants after a delay and they are told to retrieve the target words they had previously read. Tullis found that learners generate cues differently when they know that they will be used for future retrieval compared to when they are simply asked to generate descriptions. They also make their cues less distinctive and more strongly associated with the target item when they are generating cues for another person compared to when they are generating cues for themselves.
The adjustments made during cue generation might reflect underlying processes similar to those occurring during perspective-taking in language use. For instance, when taking notes in a class, the student’s goal is to jot down short reminders that will, on later reading, trigger the recollection of the larger, more complex piece of information that needs to be learned. It can be challenging to create such reminders. Often they are not successful in eliciting the appropriate recollection (e.g., when the student rereads her notes and cannot remember what they meant). In generating these cues, the student must essentially take the perspective of her future self and write notes tailored to her future self’s perspective and knowledge state. Similarly, when rereading notes, the student must use her knowledge of her past perspective to constrain the possible interpretations of the cues.

An important question for future work is whether these two seemingly analogous tasks—perspective-taking during language processing and during cue generation—do, in fact, reflect the same cognitive mechanisms, and, if so, what the underlying domain-general cognitive functions might be. For example, some work suggests that individuals who score higher on a test of working memory are better at designing referring expressions based on the perspective of their partner (Wardlow, 2013). If similar domain-general mechanisms such as working memory also confer specific benefits in effective cue generation, this would make the prediction that individuals who are successful in one domain might also be successful in the other. This has important implications for understanding the architecture of the cognitive mechanisms underlying language processing and memory. Further, better understanding how these fundamental skills are interconnected may lead to more strategic interventions for those who may have deficits in either domain (e.g., older adults).

Another important line of inquiry will be to identify how a participant’s role in a conversation, for example, as speaker, listener, or overhearer, contributes to the encoding of conversational contextual cues. Recall that Yoon and Brown-Schmidt (2013) found that while speakers showed sensitivity to the historical discourse context resulting in lexical differentiation effects, listeners did not. This finding suggests that the representation of the discourse record might differ for speakers and listeners. If so, critical questions remain about how the way in which the historical discourse is encoded in conversation—whether it is from a speaker’s perspective or a listener’s—guides how this information is used. Conversational partners are thought to encode contextually rich representations of joint experiences (Clark & Marshall, 1978), possibly through automatic memory processes that associate partners with experienced information (Horton & Gerrig, 2005), and recent findings point to a tight degree of coupling in the representations of speakers and listeners (Brown-Schmidt & Tanenhaus, 2008; Richardson,
Dale, & Kirkham, 2007). Thus, understanding the extent to which coordinated representations are normative, and when representations are not coordinated, would seem to be an important area of inquiry. Related to this question is that of the breadth of the encoding of the discourse context. For example, while it is clear that speakers maintain, and use, representations of previous referents (Van der Wege, 2009; Yoon & Brown-Schmidt, 2013), an open question is whether speakers consider other aspects of the previous discourse context as well, such as unmentioned properties of previous referents, or unmentioned objects in the discourse context. On the basis of the findings in the memory tradition, the degree to which such contextual information is encoded may depend on whether the encoding and test conditions emphasize integration of the information with the context, or instead whether associations among discussed items are more important (see Eich, 1985; Smith & Vela, 2001).

As we move closer to the goal of developing theories of conversational memory and language use, such efforts will undoubtedly benefit from increased collaboration and interaction among researchers in both memory- and language-processing traditions. Research on conversation that combines the visual world eye-tracking technique (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), well established as a key method in language processing research because it provides valuable information about the time course of comprehension processes, with methods such as explicit memory measures and signal detection analyses (Banks, 1970), will likely prove invaluable in these efforts. In addition, the emerging trend of using Amazon’s Mechanical Turk to collect data about learning and language use over time (e.g., Fine & Jaeger, 2013) will likely become an increasingly important resource as it affords collecting large amounts of data from representative populations. These new approaches, along with a more integrative view of conversational language research, one that melds language and memory perspectives (e.g., Stafford & Daly, 1984), will contribute to formulating a more unified framework of these cognitive processes.

REFERENCES


**RACHEL A. RYSKIN SHORT BIOGRAPHY**

Rachel A. Ryskin is a graduate student in the Department of Psychology at the University of Illinois at Urbana Champaign, rryskin@gmail.com.

**SI ON YOON SHORT BIOGRAPHY**

Si On Yoon is a graduate student in the Department of Psychology at the University of Illinois at Urbana Champaign, sion0912@gmail.com.

**SARAH BROWN-SCHMIDT SHORT BIOGRAPHY**

Sarah Brown-Schmidt is an Assistant Professor in the Department of Psychology at the University of Illinois at Urbana Champaign, sarahbrown-schmidt@gmail.com. Webpage: sarahbrownschmidt.com/professional.

**RELATED ESSAYS**

Theory of Mind and Behavior (*Psychology*), Amanda C. Brandone
Delusions (*Psychology*), Max Coltheart
Misinformation and How to Correct It (*Psychology*), John Cook *et al.*
Insight (*Psychology*), Brian Erickson and John Kounios
Cognitive Processes Involved in Stereotyping (*Psychology*), Susan T. Fiske and Cydney H. Dupree
Language and Thought (*Psychology*), Susan Goldin-Meadow
Concepts and Semantic Memory (*Psychology*), Barbara C. Malt
Embodied Knowledge (*Psychology*), Diane Pecher and René Zeelenberg
Gestural Communication in Nonhuman Species (*Anthropology*), Simone Pika
Attention and Perception (*Psychology*), Ronald A. Rensink
Vocal Communication in Primates (*Anthropology*), Katie E. Slocombe
How Form Constrains Function in the Human Brain (*Psychology*), Timothy D. Verstynen
Speech Perception (*Psychology*), Athena Vouloumanos
Theory of Mind (*Psychology*), Henry Wellman