CHAPTER THREE

People as Contexts in Conversation

Sarah Brown-Schmidt[1], **Si On Yoon and Rachel Anna Ryskin**
Department of Psychology, University of Illinois, Urbana-Champaign, IL, USA
[1]Corresponding author: E-mail: sarahbrownschmidt@gmail.com

Contents

Abstract

Language use in conversational settings is tailored to the knowledge and beliefs of specific conversational partners. We compare conversational partners in studies of language use to environmental context in studies of memory retrieval, and discuss the evidence of partner-specific language use with respect to the memory mechanisms involved. We outline a proposal regarding the process of encoding partner-specific contextual bindings in conversation in which we argue that formation of these bindings is limited by attention and memory processes. We discuss the way

Psychology of Learning and Motivation, Volume 62
ISSN: 0079-7421
http://dx.doi.org/10.1016/bs.plm.2014.09.003

in which this proposal accounts for the existing data in the literature, and outline a series of predictions that this view makes.

1. INTRODUCTION

A constant feature across many, if not all, domains of cognition is that cognitive processes are contextually dependent. This includes processes in the moment, such as basic perceptual processes, as well as longer term memorial representations (see Yeh & Barsalou, 2006 for discussion). For example, the perceived size of an object varies as a function of the size of the nearby objects (i.e., the Ebbinghaus illusion, see Massaro & Anderson, 1971). Likewise, judgments about the brightness of an object vary with the brightness of the nearby objects and their geometrical organization (Adelson, 1993). In decision making, preferences for one selection over another are influenced by the set of options available and the framing of the problem or choice (Huber, Payne, & Puto, 1982; Tversky & Kahneman, 1981; see Mellers, Schwartz, & Cooke, 1998 for a review). Memory is context bound as well; learners form source-item bindings, such as whom a piece of information was heard from (Johnson, Hashtroudi, & Lindsay, 1993), as well as destination-item bindings, such as whom you told a piece of information to (Gopie & MacLeod, 2009). Similarly, arbitrary environmental context affects retrieval of items from memory such that recall is improved when the learning and retrieval occur in the same location (Godden & Baddeley, 1975).

The focus of this chapter is the domain of language use and the way in which context shapes how we communicate in conversational settings. Well established in the language literature are findings that the physical context of language use guides language processing (Olson, 1970; Osgood, 1971; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). When describing a specific object, speakers typically distinguish the intended referent from other referents in the local context or referential domain (see Chambers, Tanenhaus, Eberhard, Carlson, & Filip, 2002). For example, consider a situation in which a customer is deciding which of two shirts to purchase, a checkered shirt or a striped shirt. If the customer said to the salesperson, I'll take "the shirt," that referring expression would be ambiguous. Instead, the customer must distinguish the intended referent from other potential referents in the local context, for example, by saying "I'll take the *striped* shirt." When using language, speakers use modifiers like *striped*

not only to characterize or describe the intended referent (attributive uses of modifiers) but also to distinguish the intended referent from other potential referents in the referential domain (Brennan & Clark, 1996; Osgood, 1971; Pechmann, 1989). In these examples, the language user must take note of constraints imposed by the physical environment in order to communicate effectively with others. In the present chapter, we focus on a different aspect of the context of language use, specifically the ways in which the conversational partner herself is an important source of contextual constraint. In some cases, the partner constrains the relevant physical context, for example, in situations where two conversational partners have a different perspective in the physical world.

1.1 People as Contexts in Language Use

The central goal of the present chapter is to draw parallels between context dependency in cognition broadly, and the partner specificity of language use. Here we conceptualize the conversational partner as a contextual cue in language use (also see Horton & Gerrig, 2005a) and draw parallels between the conversational partner and other types of context in language, such as objects in the visual world, in order to gain insights into how the partner shapes the use of language. In doing so, we argue that people are a particularly potent source of contextual constraint that structures and shapes the representations that underlie language processing.

Although the significance of the physical context (e.g., the physical environment and the objects therein) is well established in both memory and language (e.g., Godden & Baddeley, 1975; Osgood, 1971), here we argue that people as contexts offer a potentially more important and influential source of contextual constraint. Unlike a place or a scent, a conversational partner provides a context enriched with his or her unique perspective, knowledge, beliefs, and goals. These perspective representations and these goals guide our language, our interactions, and our memory of those interactions.

2. AUDIENCE DESIGN AND PERSPECTIVE-TAKING IN CONVERSATION

Consider that in any communicative setting, each individual brings to the situation different beliefs, knowledge, perspective, and history. Given these different backgrounds, a central problem in language use is how conversational partners coordinate representations sufficiently to be able to

effectively exchange information in a meaningful way. A classic proposal in studies of conversational language use is that conversational partners establish and grow representations of joint knowledge or common ground (Clark, 1992, 1996; Stalnaker, 1978). Common ground is thought to be formed on the basis of multiple sources of information including physically copresent information (e.g., things we can both see), culturally copresent information (e.g., information we are likely to know based on shared culture), as well as linguistically shared information (e.g., information we have talked about together) (Clark & Marshall, 1978, 1981). A central finding in studies of language use is that representations of common ground guide how we both produce and understand language (for reviews, see Brown-Schmidt & Hanna, 2011; Clark, 1996; Schober & Brennan, 2003).

2.1 Audience Design

In language *production*, common ground shapes basic choices such as what language to speak, as well as more subtle choices, such as whether to use an adjective when referring to an object. A good deal of evidence has accumulated in the literature that speakers (both children and adults) form representations of the perspective that their conversational partner holds in the physical world, and use these representations to shape how they speak (Matthews, Lieven, Theakston, & Tomasello, 2006; Nadig & Sedivy, 2002; Yoon, Koh, & Brown-Schmidt, 2012). For example, consider the situation depicted in the left panel of Figure 1. In this situation, two conversational partners are seated on opposite sides of a display with an occluder such that the person depicted at the top of the figure sees two circles,

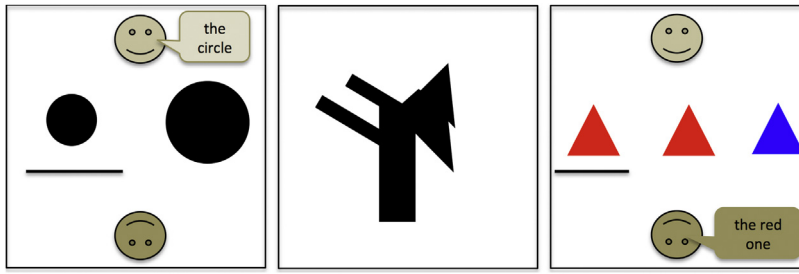

Figure 1 Example experimental displays. Left panel: Example conversational situation in which one person sees two objects and the other person only sees one object, due to an occluder. Center panel: Example unusual (tangram) image. Right panel: Example conversational situation in which one person sees three objects and the other person sees only two objects, due to an occluder.

whereas the person at the bottom sees only one circle. The environmental context (i.e., a big circle and a small circle) and the partner-specific context (i.e., one mutually visible circle) are at odds. In situations such as this one, if the person at the top of the display wished to ask his or her partner to touch the larger of the two circles, a request such as "Please touch the circle" would suffice. In this case, the unmodified expression "the circle" is sufficient to identify the intended referent. From the speaker's perspective, the intended referent would be better described as "the *large* circle," whereas from the listener's perspective, the adjective "large" is unnecessary, and there is some evidence that such unnecessary adjectives may be confusing to listeners (Engelhardt, Bailey, & Ferreira, 2006; Gann & Barr, 2012). In situations such as this one, both adults and 5- to 6-year-old children are successful at designing expressions from the addressee's perspective, e.g., saying "the circle," rather than "the large circle" about 50% of the time (Nadig & Sedivy, 2002). By contrast, when both the speaker and addressee see the two items in the size-contrasting set (e.g., both circles), adjectives are used between 70% and 100% of the time, depending on the complexity of the display and other such factors (Brown-Schmidt & Konopka, 2011; Heller & Chambers, 2014; Brown-Schmidt & Tanenhaus, 2006; Nadig & Sedivy, 2002; Sedivy, 2005). The fact that adjective use is significantly less frequent when an adjective is unnecessary from the listener's perspective shows that speakers encode the listener's perspective, and that their perspective can guide language production. To be sure, this audience design process is not perfect, and speakers do not always produce utterances that are tailored to the addressee's perspective (Horton & Keysar, 1996; Lockridge & Brennan, 2002; Wardlow-Lane, Groisman, & Ferreira, 2006). However, even in cases where speakers are under time pressure and their own egocentric perspective conflicts with that of the addressee, they are generally successful at adopting the addressee's perspective more than half of the time (e.g., Horton & Keysar, 1996).

Shared knowledge from past linguistic experiences guides language production as well. For example, in a classic referential communication task (Krauss & Weinheimer, 1966), one partner, the Director, gives another partner, the Matcher, instructions on how to rearrange a set of abstract "tangram" images. The partners work together to rearrange the images multiple times in different orders. During the course of this process, the partners typically develop shared names for the images; these names become shorter and more opaque over time (Clark & Wilkes-Gibbs, 1986). For example, given the image in the center panel of Figure 1, the way the Director refers

to that image might develop across trials to a short and concise label that both partners would easily understand; see hypothetical example utterances from the first three trials in the task below:

Trial 1. *Director*: uh, it looks like a person with the arms out to the left and head back kind of like they are dancing. *Matcher 1*: Uh, arms out, OK.
Trial 2. *Director*: the person with the arms out that is dancing. *Matcher 1*: got it.
Trial 3. *Director*: the dancer. *Matcher 1*: mm-hmm.
Trial 4. *Director*: (instructing a new Matcher) uh, it looks like a dancer with the arms out to the left and the head tilted back to the right. *Matcher 2*: dancer with arms, OK.

Critically, this naming process is partner specific, and tailored to the knowledge of the partner. After the Director and Matcher establish names for a set of unusual images as in Trials 1–3, if the speaker describes the images for a new partner, they consistently use longer referring expressions that contain more new content, as in Trial 4 (Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Heller, Gorman, & Tanenhaus, 2012; Horton & Gerrig, 2002; Horton & Spieler, 2007; Wilkes-Gibbs & Clark, 1992). These expanded descriptions, when speaking with a new partner, are generally thought to be necessary in order for the new, naïve partner to understand the referential labels. Indeed, some research shows that if an overhearer listens in on such conversations, they have difficulty interpreting labels such as "the dancer" in example Trial 3 (Schober & Clark, 1989), consistent with the idea that these collaboratively established labels are partner specific and opaque to individuals not involved in the original conversation.

An emerging domain for future work, and one focus of the present chapter, is how interlocutors form memorial representations of what different partners do and do not know, and how these memory representations guide and constrain audience design processes. One clear finding in this literature comes from Horton and Gerrig (2005b) who demonstrated that speakers were more likely to appropriately design expressions with respect to the perspective of a naive addressee (like Matcher 2, above) in situations where the knowledge shared with each addressee in the situation was distinctive vs nondistinctive (also see Horton & Slaten, 2012). This finding suggests that distinctiveness of the information associated with different partners may help to learn partner-specific information (cf. Heller, Gorman, & Tanenhaus, et al., 2012; Gorman et al., 2013). Findings such as this one emphasize the importance of understanding the memory representations that support partner specificity of language use.

2.2 Perspective-Taking

As in language production, during language *comprehension*, common ground shapes both the "online" (immediate) processing of language as well as the ultimate understanding of what the conversational partner has to say. Much of the relevant literature focuses on the physical context of language use. When processing the speech of another person, knowledge of what the speaker can and cannot physically see guides considerations of the speaker's intended meaning (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Ferguson & Breheny, 2012; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008). For example, consider the situation depicted in the right panel of Figure 1. In this situation, the person depicted at the top of the panel sees three triangles, whereas the person at the bottom of the panel only sees two. If the person at the bottom of the panel were to give an instruction to his or her partner, "Pick up the blue triangle and put it on *the red one*," the instruction is ambiguous and potentially confusing from the addressee's perspective because the addressee sees two different red triangles. However, if the addressee could take the speaker's perspective into account, the triangle that is hidden from the speaker could be ruled out as a potential referent, and the speaker could be understood to mean the triangle in the middle of the display.

Hanna et al. (2003) examined situations such as this one in which a speaker gave an instruction to an addressee which contained a referential ambiguity (e.g., "the red one" matches more than one potential referent). Critically, this ambiguity could be resolved if the addressee took into account which items were in common ground. Their experiment drew on previous work in the visual world paradigm (Tanenhaus et al., 1995), in which eye movements during the moment-by-moment interpretation of language are used to understand the addressee's candidate interpretations of the unfolding sentence. Analysis of addressee eye gaze during interpretation of the ambiguous expression revealed that addressees readily interpreted the ambiguous expression "the red one" as meaning the red triangle in common ground. This result shows that addressees formed a representation of their partner's physical perspective on the scene, and used this perspective representation to guide the interpretation of the speaker's utterances. Importantly, however, addressees did fixate the privileged ground triangle (the one that only the addressee could see) more than chance, showing that, as in language production, perspective information is not a complete constraint on language processing.

A growing body of evidence shows that knowledge about shared linguistic experience similarly acts as a significant, partial constraint on language understanding (Brennan & Hanna, 2009; Brown-Schmidt, 2009a, 2009b, 2012; Horton & Slaten, 2012; Metzing & Brennan, 2003; Richardson, Dale, & Kirkham, 2007; Richardson, Dale, & Tomlinson, 2009). For example, consider that if someone asks you a question such as "What's for dinner," it is reasonable to assume that the speaker does not, in fact, know what is for dinner, but that he or she believes that you do know. Based on this observation that speakers typically ask informational questions when they do not know the answer, Brown-Schmidt et al. (2008) designed an experiment to examine whether these perspective representations guide online language comprehension. They examined situations in which a speaker asked an addressee questions such as "What's below the cow that's wearing the hat?" in the context of a collaborative game in which some of the game pieces were seen by both the speaker and the addressee (common ground) and others were only seen by the addressee (privileged ground). The question ("What's below the cow…") was temporarily ambiguous because there were two different cows on the game board, e.g., one wearing a hat, and the other wearing shoes. To examine the influence of shared linguistic experience, Brown-Schmidt and colleagues created situations in which, just prior to the critical question, the conversational partners discussed the animal below the other cow (the one wearing shoes), thus bringing that game piece into common ground. In a control condition, they discussed a different, unrelated animal. When the animal below the cow with shoes was already common knowledge, when interpreting the critical question, e.g., "What's below the cow…," addressees quickly interpreted the question as asking about the *other* cow (the cow with the hat), as, after all, the speaker already knew what was below the cow with shoes. By contrast, in the control condition, the speaker did not know what was below either cow, and addressees considered both cows to be potential referents until the disambiguating word "hat." These findings show that addressees form representations of what their partner does and does not know based on both the discourse history and the physical context, and use these representations to guide the online processing of language.

Taken together, this literature shows that in conversation we form representations of the perspective of our conversational partners, and use these representations to guide language use and processing. These findings further show that it is not simply the physical world that guides language use. Instead it is our construal of the physical world, and our beliefs about others'

construal of that world that guide how we converse. In what follows, we discuss three special cases of audience design and perspective-taking in conversational settings in order to begin to ask questions about the breadth and scope of the role of perspective representations in language use. These cases are: audience design in multiparty conversation (Section 2.3), conversational goals and perspective-taking (Section 2.4), and spatial perspective-taking (Section 2.5).

2.3 Audience Design in Multiparty Conversation

In the previous sections, we focused primarily on situations in which individuals adjust to the knowledge, beliefs, and goals of a single other person. On the whole, this research suggests that interlocutors are generally successful at adjusting to the perspective of a single partner in a dialog. Less clear, however, are the mechanisms involved in this process. In this section, we use multiparty conversation as a test case to examine how audience design processes in dyadic conversation scale up to conversation among triads in which the three partners share different amounts of common ground. In doing so, we begin to address the mechanisms by which conversational partners form representations of the perspectives of others, and apply these perspective representations to the process of utterance design.

Consider that in many situations, three or more individuals interact and each brings to the situation his or her own distinct perspective. In dialog situations, fairly simple definitions of common ground (information both partners jointly know) and privileged ground (information only one partner knows) offer a good starting point for understanding how a speaker and listener are likely to understand one another. However, the situation is considerably more complex in conversations with three or more individuals because distinct common ground is likely to be shared among the different dyads within the larger group. For example, consider a situation in which Duane, Otto, and James are chatting with one another at a birthday party and there is a large present sitting in the center of the room. Imagine that earlier in the day, Duane revealed to Otto that the present is in fact, a bicycle. In this situation, Duane and Otto share common ground for the identity of the present, but James does not. If Duane were later to discuss the present with both Otto and James in a three-party conversation, their inconsistent knowledge poses a conundrum: should Duane speak with respect to common ground shared with Otto, or the lack of common ground held with James? While such situations are common, surprisingly little experimental work has addressed how common ground is managed in such situations.

Yoon and Brown-Schmidt (2014a) examined three-party situations such as this one, in which different dyads within the larger group held distinct common ground. The first goal of this research was to evaluate whether conversational partners are able to simultaneously maintain multiple, distinct representations of the common ground held with different individuals within the larger group. The second goal was to begin to understand how these representations of perspective are integrated into language production and language comprehension. In the first phase of the experiment, a Director and a Matcher completed several rounds of a referential communication task in order to develop shared names for a set of tangram images (see left panel of Figure 2). Following this first phase of the task, we compared two different conversational situations. In the *dialog* condition, the Director continued to describe the tangram images for the original Matcher; thus during both the first and second phases of the task, the same partners with the same common ground conversed. By contrast, in the *three-party* condition, a third, naïve partner joined the conversation (right panel of Figure 2). In this condition, the Director and the original Matcher maintained their common ground for the tangram labels (e.g., "the dancer"), but the new Matcher did not share this knowledge.

The critical analyses compared how the Director described the images when addressing only Matcher 1, who was knowledgeable about the image labels, with descriptions of the same images when simultaneously addressing Matcher 1 and naïve Matcher 2. Yoon and Brown-Schmidt found that Directors produced longer and more disfluent referential expressions in

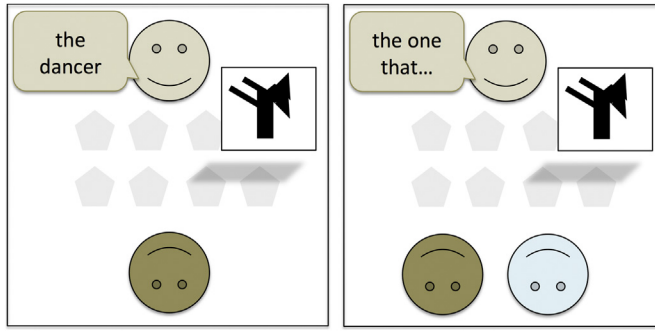

Figure 2 Schematic of a multiparty conversation. Left panel: The Director and Matcher 1 establish shared names for a series of images. Right panel: The Director describes images for both the knowledgeable Matcher 1 and a new partner, Matcher 2, who does not know the previously established image labels.

the three-party conversation where the naïve listener was present. Thus, similar to situations in which a knowledgeable speaker talks to a single naïve addressee (Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992), speakers adapted to the knowledge of the less knowledgeable addressee when simultaneously addressing a knowledgeable and a naïve listener. This process of audience design was reflected both in the inclusion of new content words (e.g., "the dancer with *arms sticking out*"), and also in the increased disfluency rate (e.g., "*thee uh*…dancer…"), which likely reflects the extra effort that speakers put in to the process of reformulating their descriptions (see Clark & Wasow, 1988; Ferreira, 1991). Thus in this situation, speakers chose to weigh the needs of the naïve addressee over the needs of the knowledgeable addressee, possibly in order to make sure that both partners were able to comprehend the instructions. Note that by doing so, Directors sacrificed to some extent the speed and efficiency with which the knowledgeable Matcher 1 likely could have interpreted the shorter, previously established expressions.

In a second experiment, Yoon and Brown-Schmidt (2014a) examined whether an addressee such as Matcher 1, is aware of the speaker's need to adjust referential descriptions due to the naiveté of Matcher 2. If so, when interpreting a lengthy, disfluent instruction in a situation such as the one depicted in the right panel of Figure 2, Matcher 1 may be able to take this into account in order to facilitate interpretation of the speaker's sentence. In this experiment, the Director and Matcher 1 first established names for a large set of abstract images as before. Then, at test, the Director and Matcher viewed a scene that contained three old images with established referential labels and one novel image. The Director was a laboratory assistant who gave the Matcher an instruction to click on one of the four images; on critical trials the Director always referred to one of the old images with a label that had been established in the first part of the experiment. The first critical manipulation was whether the Director was addressing just Matcher 1 or both Matcher 1 and naïve Matcher 2. The second manipulation was whether the experimenter referred to the image using the fluent, established expression, e.g., "the dancer," or a lengthy disfluent description, e.g., "the one that looks like…a dancer." Yoon and Brown-Schmidt tracked the eye fixations that Matcher 1 made to the four objects in the display in order to monitor their interpretation of these referential descriptions. Based on previous research examining the interpretation of disfluent instructions (Arnold, Hudson Kam, & Tanenhaus, 2007; Arnold, Tanenhaus, Altmann, & Fagnano, 2004; Barr & Seyfeddinipur, 2010), we expected that in dialogs

between the Director and Matcher 1, the disfluency would initially be interpreted as referencing the new object, which did not have a previously established label. This bias to interpret disfluent expressions as referring to new objects is thought to reflect an inference that the disfluency is due to the speaker effortfully trying to come up with an appropriate referential label. The critical question, then, is whether this disfluency-new object bias would be attenuated in situations where Matcher 1 could attribute the disfluency to the Director's need to redesign the instruction to accommodate Matcher 2′s naiveté, rather than a signal that the Director is referencing a novel object.

The analysis of Matcher 1's eye fixations revealed that fluent referential labels were readily interpreted, regardless of the presence of the new addressee. This result replicates previous findings from dialog situations (Brown-Schmidt, 2009a; Metzing & Brennan, 2003) that established referential labels are rapidly interpreted, and shows that the presence of the naïve addressee did not disrupt the interpretation process. A different pattern of results emerged during the interpretation of disfluent instructions. Here, identification of the intended referent (e.g., the dancer) was overall slower than with fluent instructions, which is to be expected considering that these instructions were much longer (compare "the dancer" with "thee uh…dancer…"). More importantly, consistent with the hypothesis that disfluency can be attributed to the presence of the naïve matcher, rather than the Director's intention to reference a new object, the eye-tracked Matcher was significantly more likely to fixate the target object (e.g., the dancer) during disfluent instructions when the Director was simultaneously addressing both Matcher 1 and Matcher 2, vs situations in which the Director was only addressing Matcher 1. This result shows that even when the speaker and addressee have established referential labels for these images, addressees can cancel expectations for these established labels, given sufficiently motivating contextual changes, such as a new addressee who would be confused by an opaque label that he or she had not previously been exposed to.

In sum, the results of these experiments reveal a surprising degree of flexibility in the ability of conversational partners to adapt to the presence of a third party in the conversation. These findings support the idea that conversational partners can form and hold onto multiple distinct representations of common ground held with different individuals within the larger situation. A central open question, and focus of research in our laboratory, concerns the processes by which these representations are brought to bear on the process of language use in conversational situations. For example, are there limits to the number of people for whom we can maintain distinct perspective

representations in multiparty conversation? Likewise, what principles govern audience design in situations where accommodating one addressee's perspective dramatically impairs another addressee's ability to understand?

2.4 Conversational Goals and Perspective-Taking

Earlier we compared the conversational partner to a type of contextual cue and asserted that people may serve as a particularly potent source of contextual constraint. One feature of conversational partners that makes them special is that they posses a unique perspective in any situation, and speakers and listeners alike are sensitive to these perspective representations. Another feature of conversational partners that distinguishes them from other types of contextual cues is that each person brings to the conversation a set of goals for what they want to get out of the situation. Different types of utterances can be used to achieve different types of goals, such as to make a statement, ask a question, give a command, or express a desire (Searle, 1969). Yoon et al. (2012) tested the hypothesis that conversational goals influence perspective-taking processes. Yoon et al. compared two types of goals that speakers might assume during language production, specifically the act of informing vs the act of requesting. Yoon et al. hypothesized that different types of conversational goals make distinct demands on perspective-taking. Specifically, they reasoned that when informing someone of a new piece of information, the speaker may not need to pay close attention to the addressee's perspective in the situation, compared to a case in which the speaker wishes to make a request of the addressee. When a speaker is requesting the listener's help, the speaker may be particularly motivated to ensure that the addressee understands the speaker's intention, in order to make sure the speaker gets what he or she wants. Paradoxically then, fulfilling an egocentric desire to get someone to do something for you, may force you to pay close attention to that person's perspective.

Yoon et al. (2012) evaluated this hypothesis by examining whether goals (informing vs requesting) modulate the process of audience design. On critical trials, a to-be-described display included two objects of the same type that differed only in size (e.g., a large cup and a small cup, similar to the setup in the left panel of Figure 1). The experimental display was set up such that the speaker and addressee both saw a large cup, making that object part of their common ground. At the same time, the speaker, but not the addressee, additionally saw a small cup, making that object in the speaker's privileged ground. On critical trials, the speaker referred to the common ground cup.

Based on previous findings (Nadig & Sedivy, 2002), if speakers take into consideration the addressee's perspective, in these situations they should be more likely to refer to the common ground cup without using a modifier, as in "the cup." The critical manipulation was whether the speaker made a request of the listener, or informed the listener. In the request condition, speakers asked the addressee, e.g., "Can you move the (large) cup to the right?"[1] In the inform condition, the speaker informed the addressee of an action that a nearby experimenter was about to perform, as in "The experimenter will move the (large) cup to the right." Analysis of these utterances showed that speakers were significantly less likely to use a modifier that was unnecessary from the addressee's perspective (e.g., large) when requesting than when informing (37% vs 60%). There was no difference in modification rates on control trials where the speaker and the addressee shared the same perspective. Thus, speakers were more likely to consider their partners' perspective when making a request.

In summary, while common ground is fundamental to coordinating language use in conversation (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986), these findings show that the degree to which speakers adhere to common ground may depend on their own goals. Under at least one theoretical perspective, language processing is thought to involve the simultaneous combination of multiple probabilistic sources of information or constraint (e.g., MacDonald, 1994; Tanenhaus & Trueswell, 1995). We propose that the weighting of these constraints is modulated by one's goals for engaging in the conversational situation in the first place (also see Buxó-Lugo, Toscano, & Watson, 2013; Yee & Heller, 2012). A key goal for future research is to understand the scope of such goal-based modulations of sentence processing, and to identify whether there are sources of information that are routinely used to guide language processing regardless of the goal state.

2.5 Spatial Perspective-Taking

Another way in which the conversational partner constitutes a unique source of contextual constraint is through their place in, and interaction with, the physical environment. In face-to-face conversation, our representation of the other person's perspective is constrained by their spatial

[1] Note that the experiment was carried out in South Korea and the example utterances are translational equivalents of the Korean sentences that the participants produced.

orientation and our expectations about how any given person might act on the world. For instance, when addressees hear a request like "Hand me the cake mix…," they typically interpret it as a request for a box of cake mix that is out of the speaker's own reach, but one that the addressee could reach from her own position (Hanna & Tanenhaus, 2004). Thus, information about the range of the speaker's abilities to interact with the physical environment is spontaneously encoded and used to facilitate language comprehension. These effects go beyond tracking which potential referents in a visual scene are and are not visually or physically available to a conversational partner (Hanna et al., 2003; Heller et al., 2008; Nadig & Sedivy, 2002), to include encoding of other aspects of being immersed in the visual world, such as spatial viewpoint. Thus, in the same way that each person brings to the conversation a different base of knowledge and beliefs, in any dialog, the speaker's and listener's representation of the spatial relations among objects that surround them differ. For example, in the left panel of Figure 3, two dining partners are discussing the appetizers that they ordered but do not remember the names of. For diner A, the spinach (the green plate) is on the left but for diner B, the spinach is on the right. In order to avoid an awkward interaction, if diner A wants to try the spinach, he might choose to adjust to B's spatial perspective and say, "Pass me the dish that's to the right." It is this spatial relational information that we focus on in this section.

Schober (1993) examined how conversation partners coordinate their spatial language when describing the locations of mutually visible objects

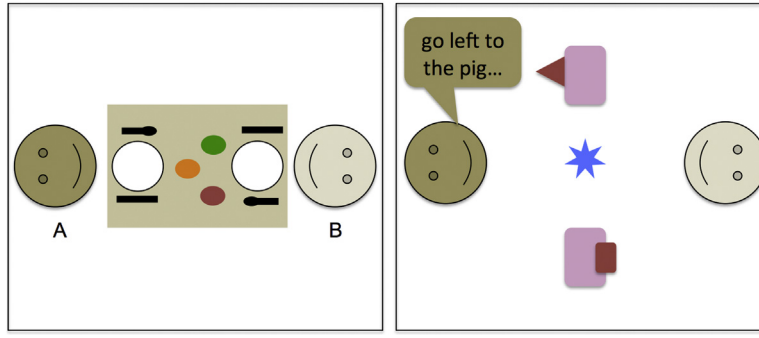

Figure 3 Example conversational situation in which the partners have different spatial viewpoints (180° transformation) on the scene. Left panel: Partners discuss spatial layout of plates on a dinner table. Right panel: Schematic of experimental display used in Ryskin et al. (2014) in which the addressee follows instructions to drag a star about the screen.

relative to each other. He found that speakers and listeners spontaneously agree on which perspective is to be used throughout the course of the dialog. They rarely, if ever, explicitly announce which viewpoint they will be operating under (e.g., "When I say left I mean *my left*"). Yet, pairs entrained on spatial perspectives in the same way that conversation partners entrain on idiosyncratic referential labels (Brennan & Clark, 1996; Yoon & Brown-Schmidt, 2013). For example, if the director started out by using the addressee's perspective (e.g., "uh it's the one on the right" meaning the addressee's right), clarification about which perspective was being used was rarely needed and the dyad would then continue using this implicitly agreed upon perspective for the remainder of the interaction (see also Garrod & Anderson, 1987). Once a reference frame for the spatial descriptions was jointly established, speakers rarely switched to a different one.

In much the same way that speakers are sensitive to the visual copresence of objects when designing referring expressions (Nadig & Sedivy, 2002), speakers choose which spatial perspective to use while formulating their utterances based on pragmatic information about the listener. For instance, when told to give instructions to an imaginary listener who will be positioned at another location, speakers often formulate spatial instructions (e.g., place the X in the upper-right-hand corner) from the perspective of this imagined recipient. On the other hand, when the listener is physically present, speakers produce more egocentric instructions, presumably with the assumption that the present listener will request a clarification when the instruction is unclear, while the imagined listener cannot. The opportunity for conversational feedback is thought to provide necessary opportunity for coordinating meaning in conversation (Bangerter & Clark, 2003; Clark & Krych, 2004), and may partially explain why speakers do not design utterances from the addressee's physical perspective 100% of the time (e.g., Nadig & Sedivy, 2002). When partners are mismatched in terms of their spatial reasoning abilities, high-spatial-ability speakers elaborate more on their instructions when paired with a low-spatial-ability listener, compared to when they are matched with an equally adept listener (Schober, 2009), despite having no explicit knowledge of their respective spatial skills.[2] This adaptability suggests that speakers and listeners encode spatial contextual information beyond an interlocutor's spatial orientation. They store a

[2] Experts in a particular semantic domain (e.g., knowledge of New York City) similarly adjust their language when talking about that domain to a naïve person, though in these cases speakers are aware of their own knowledge (Bromme, Jucks, & Wagner, 2005; Isaacs & Clark, 1987).

much more elaborate representation that includes details such as the perspectives that have been jointly established previously and even the level of success with which the interlocutor was able to make use of certain spatial information.

Further evidence for the rich representation of an interlocutor's spatial viewpoint comes from findings that listeners are quick to adjust to a spatial perspective that differs from their own. Ryskin, Brown-Schmidt, Canseco-Gonzalez, Yiu, and Nguyen (2014) found that participants are able to use information about a conversational partner's spatial perspective during online interpretation of spatial language. In a visual world eye tracking paradigm, participants heard instructions to drag objects around the screen (see Figure 3, right panel), such as to pick up a star (with the computer mouse) and "Go left to *the pig with the hat*." These instructions were given either from the participant's egocentric perspective (i.e., "left" = participant's left) or the opposite perspective (a 180° rotation; "left" = participant's right). Prior to each block of trials, participants were explicitly instructed, through the use of a visual diagram, which viewpoint the speaker would be using for the upcoming set of trials (i.e., 0° rotation or 180° rotation). The scenes were designed such that the critical instructions were temporarily ambiguous between two potential referents. For example, the italicized noun phrase, e.g., "*the pig with the…*" was temporarily consistent with two different pigs on the screen, one of which was located to the *left* of the starting position (e.g., a pig wearing a hat), and the other was located to the *right* of the starting position (e.g., a pig wearing a purse). Critically, this temporary ambiguity could be resolved early if the participant integrated the speaker's perspective into the interpretation of the sentence. Analysis of participant eye movements as they interpreted the temporarily ambiguous referring expression, "the pig with the…," revealed that instructions that were generated from the opposite spatial perspective posed challenges and delayed processing. However, despite these challenges, participants showed a clear target bias well before the onset of the disambiguating word (e.g., hat), showing that even when spatial perspectives are misaligned, listeners are able to use knowledge about the speaker's adopted spatial viewpoint to understand their meaning. In sum, addressees are able to rapidly retrieve the relevant spatial perspective information and use it to constrain the interpretation of a sentence.

A question remains as to whether the memory representations created by speakers and listeners of each other's perspectives are enduring over time. Listeners may tackle the task of taking the speaker's spatial perspective in one of two ways. They might (1) approach each spoken utterance

independently and perform a spatial transformation of their own perspective while interpreting the utterance or (2) store the speaker's perspective and interpret all subsequent utterances from that remembered viewpoint. Because the speaker's viewpoint is likely to be relatively stable during a conversation, listeners can predict, with some certainty, that their conversational partner will continue using the same perspective throughout the conversation. If so, it may be computationally efficient to store memories of an interlocutor's perspective rather than undergoing a mental translation every time they begin to speak. If perspectives are stored, we would expect to see evidence of a switch cost when conversational partners switch which spatial perspective they are speaking from (also see Gratton, Coles, & Donchin, 1992). Some evidence for this computational burden comes from the fact that, when speakers switch between perspectives, listeners experience a cost in reaction times: Ryskin et al. manipulated whether participants heard multiple trials in a row with the same perspective (either their own or the opposite viewpoint) or switched from one perspective to the other as they moved to the next trial (i.e., from their own to the opposite view, or from the opposite view to their own perspective). Ryskin et al. found that interpretation of the spatial term (e.g., Go left…) was slower when the instructions on the previous trial had been from a different perspective. This decrement occurred even when participants switched into their own perspective, suggesting that the representation of the opposite viewpoint had been learned well enough to interfere with the egocentric viewpoint. Thus, in much the same way that interlocutors learn partner-specific labels for objects in conversation (Metzing & Brennan, 2003), these findings show that interlocutors similarly learn and store representations of the partner's spatial viewpoint (also see Galati, Michael, Mello, Greenauer, & Avraamides, 2013).

2.6 Interim Summary

Research on perspective-taking in language processing shows that we form representations of the perspective of other individuals, and use these representations to guide both language production and language comprehension. These representations are formed for multiple conversational partners (Yoon & Brown-Schmidt, 2014a), and include information about visual and spatial perspective in the world (Hanna et al., 2003; Schober, 1993), the discourse history (Brennan & Clark, 1996), and conversational goals (Yoon et al., 2012). Much of this literature grew out of theoretical concerns regarding whether occasional errors in perspective-taking reflect egocentric default

processes (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003), or instead are evidence of the immediate and probabilistic combination of multiple constraints (Hanna et al., 2003). With a large literature showing clear effects of perspective on language processing, it is now generally thought that perspective does guide language processing. The focus of much of the ongoing research and theoretical debate now concerns the mechanisms and time course by which perspective is incorporated into language processing (Barr, 2008; Brown-Schmidt & Hanna, 2011), as well as the memorial processes that support representation of perspective (Horton & Gerrig, 2005a). Proposals regarding the nature of perspective representations in memory include rich, diarylike representations of joint experience (Clark & Marshall, 1978), simple associations between people and concepts (Horton, 2007), as well as the idea that audience design may be guided by one-bit cues, e.g., as to whether or not a conversational partner sees a particular object (Galati & Brennan, 2010). In what follows, we set aside questions of time course and explore in some detail the mechanisms by which we form representations of the perspective of the conversational partner.

3. PEOPLE AS CONTEXTS IN CONVERSATION: MECHANISMS OF ENCODING

Given a broad range of findings that speakers and listeners tailor their language use to the perspective of specific conversational partners (Brennan & Clark, 1996; Metzing & Brennan, 2003; Wilkes-Gibbs & Clark, 1992), we now turn to focus on the cognitive mechanisms which support these processes. What are the relevant mechanisms required for the speaker to produce an utterance such as "the dancer," tailored to the knowledge of the addressee, and for the addressee in turn to understand the speaker's perspective when interpreting the same expression?

Answering this question may be informed by related findings of partner-specific processing outside the domain of perspective per se. Conversational partners not only form representations of the knowledge and perspective of their partner but also form representations of the relationship between individuals or words and information in a variety of domains including: talker–word pairings (Creel & Tumlin, 2011; Creel, Aslin, & Tanenhaus, 2008), talker-verb argument structure pairings (Kamide, 2012), talkers and their preferences (Creel, 2012, 2014), words and their prosodic or acoustic properties (Bradlow, Nygaard, & Pisoni, 1999; Goldinger, 1998; Magnuson & Nusbaum, 2007), words and background noise (Creel, Aslin, & Tanenhaus,

2012), words and properties of room such as an odor (see Parker, Dagnall, & Coyle, 2007), words and indexical properties of the speaker such as talker accent (Dahan, Drucker, & Scarborough, 2008), and talkers and their social category (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008).

Given that this wealth of contextual information is used, it becomes important to understand the cognitive pressures that shape how this information is encoded in conversational settings. In one of the earliest discussions of the mechanisms that support formation of common ground, Clark and Marshall (1978, 1981) proposed that conversational partners assume mutual knowledge for a given piece of information when both their partner and the information are copresent (visually, auditorily, culturally, etc.) and both partners pay attention to the information. For example, if both partners look at a picture of a duck in a given situation (and notice each other looking), they could be reasonably sure that the duck is in their common ground. On Clark and Marshall's view, this event that established the duck as common ground is stored in memory in rich, diarylike representations ("reference diaries"). In turn, these representations support audience design and perspective-taking in subsequent language use.

Interest in the memory contributions to audience design in language production has increased in the past few years, following a series of papers by Horton and Gerrig (2005a,b; Horton, 2007). Horton and Gerrig (2005a) proposed that the assessment of common ground and the use of this information in language production are governed by basic associative memory processes. When we experience a piece of information with another person, episodic memory traces are formed, resulting in associations between the information and that person in memory. The conversational partner later functions as a contextual memory cue that supports access to the relevant traces. Unlike Clark and Marshall's (1978) proposal, as long as the information is associated with a person, that information may be treated as common ground, regardless of whether the information is, in fact, jointly known to both partners. On this association-based view, weaker associations may have less of an effect on subsequent language use compared to stronger associations. Further, associations should lead to telltale errors in assumed common ground when associations are strong, but common ground is not in fact held.

Here we take a somewhat different approach, focusing on the encoding of partner-specific contextual information during conversational situations, rather than the retrieval and use of this information. We take as a starting point Horton and Gerrig's (2005a,b; Horton, 2007) suggestion that associations between partners and information are formed in conversational

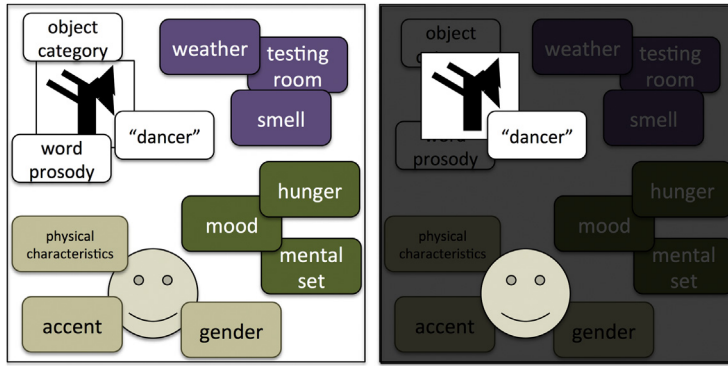

Figure 4 Schematic of relations hypothesized to be encoded in memory in the unbounded case (left panel) and the bounded case (right panel). In the bounded case, only a small subset of possible relations are encoded.

settings. Here we focus more specifically on the processes that govern the encoding of partner-specific bindings, in order to understand what bindings are likely to be learned and thus be positioned to guide future language use. In doing so, we develop a proposal that draws on findings from the language literature, as well as insights from the attention and memory literature. We argue that there are key limits on the partner-specific contextual representations that are likely to be formed.

3.1 Motivation for Limits on the Formation of Partner-Specific Representations

We propose that the formation of partner-specific representations during conversation is limited by attentional, memorial, and communicative constraints. If so, this predicts that in a given communicative situation, only a limited number of contextual bindings (e.g., partner-word bindings) should be encoded (right panel of Figure 4). Alternatively, if all contextual associations are formed automatically, regardless of whether they are communicatively relevant, we ought to expect a rather large and possibly unbounded number of associations to be formed (left panel of Figure 4). The unbounded case might include large numbers of both partner-specific bindings (e.g., talker-word bindings, Creel et al., 2008), and bindings that do not involve the partner, such as word-room bindings, Parker et al., 2007).[3]

[3] Note that we do not intend to equate Horton and Gerrig's automatic associations view (Horton, 2007) with this unbounded associations case, as their proposal focused primarily on those associations formed with conversational partners.

In what follows, we argue that existing data from the conversational literature are more consistent with the idea that the formation of partner-specific bindings is limited, in particular by the communicative relevance of the conversational partner as an individual (cf. Horton, 2007). We also outline predictions of this proposal, many of which are currently being evaluated in our laboratory.

3.2 Attentional and Memorial Constraints on Learning

One reason to think that people may serve as contexts in language only to the extent that they are meaningfully related and relevant is that it may be computationally infeasible to encode all possible contextual relations between the language used in a conversational setting and the contextual cues in the situation.

Motivation for why there may be significant limits on the degree to which partner-specific representations are formed during conversation comes from findings in the memory literature regarding the circumstances that modulate the size of the environmental context effect. Classic work shows a memory recall benefit when study and test are in the same environmental context (Godden & Baddeley, 1975), whereas the effect is not always observed with recognition memory tests in the laboratory (where contexts are, e.g., distinct laboratory rooms; Smith, Glenberg, & Bjork, 1978). Even with extreme environmental cues, context effects are not always observed in recognition memory (i.e., under water vs on land, Godden & Baddeley, 1980). The environmental context effect in recall can be diminished if participants in a novel room are asked to mentally reinstate the original context (Smith, 1979), suggesting that the effect may be related to memory retrieval strategies (also see Parker et al., 2007). Mulligan (2011) observes the environmental context effect in an explicit test of memory (cued category recall) but in an implicit test (category production), context effects are only observed for the subset of subjects who report, in a postexperimental questionnaire, that they noticed that the study and test phases were related. Similarly, Eich (1985) examines environmental context effects (again, different laboratory rooms) in word recall and finds the effect on memory recall is obtained only when participants are actively encouraged to integrate the training room with the trained stimuli (they were asked to imagine an image of each word situated in the room), but not when simply asked to imagine the word in isolation. These results are of interest because they suggest that strategic factors may be in play.

In the case of partner-context cues in conversational situations, then, these findings offer two predictions. First, the more integral the partner is to the communicative situation, the more likely the partner as context is likely to be bound to the studied information (e.g., the label "the dancer" for an unusual tangram image). As a result, when partner-specific knowledge is acquired in communicative contexts, it should be more likely to affect subsequent language use compared to partner-specific knowledge acquired in a noncommunicative context. Second, whether we do or do not observe partner-specific effects on subsequent language use may be under the metacognitive control of conversational participants. In situations where an individual mentally reinstates (Smith, 1979) a previous event that established common ground, audience design and perspective-taking may be more likely.

3.2.1 *Why Might the* Number *of Behaviorally Relevant Associations Be Limited?*

Consider that in any given communicative situation an unbounded number of potential bindings between item and context could be formed (e.g., see Figure 4, left panel). While it is clear that children and adults use, among other things, co-occurrence statistics to form object-label bindings (Smith, Suanda, & Yu, 2014), the third-level bindings are too numerous to fully consider: person-object-label, room-object-label, scent-object-label, weather-object-label, etc. Given limits on visual attention and working memory (e.g., Engle, 2002; Simons & Chabris, 1999), there must be limits on the number of such relations that can be encoded in a given situation (see Doumas, Hummel, & Sandhofer, 2008; Saiki, 2003). According to one proposal, limits on the focus of attention allow only one relation to be encoded at a time, such that a glance between a square and a circle would result in encoding that the circle is to the left of the square but not the reverse relation—that the square is to the right of the circle (Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012; Roth & Franconeri, 2012). Thus, while most of the time we experience the world in rich detail, in reality, at any one point in time, we are only perceiving and storing a small amount of the information available in the world around us.

Extending these arguments to conversational situations, basic attention limits may constrain the number of relations that can be encoded in any given communicative situation (for various ideas about the way in which language users solve the problem of attentional limits in discourse understanding, see Walker, 1996; Fletcher, Hummel, & Marsolek, 1990; Grosz & Sidner, 1986). If so, this potentially offers an attention-based explanation

for why an overhearer in a conversation is not assumed to have full knowledge of the contents of the conversation. For example, if partners establish joint labels for a tangram image, e.g., "the dancer" in a situation where an overhearer is listening in, if the speaker subsequently turns to address the overhearer, the speaker does not assume he or she knows the established labels, and instead uses longer, more elaborated expressions (Wilkes-Gibbs & Clark, 1992). This phenomenon has been typically explained in terms of the lack of established common ground between the speaker and overhearer, consistent with findings that overhearers, in fact, do not learn these names as well as true conversational partners (Schober & Clark, 1989). However, the locus of the effect may be much simpler than an inference that the overhearer lacks common ground. Instead, attentional limits may have prevented the speaker from binding the overhearer to the established referential terms during the conversation in the first place. If this account is correct, one clear prediction that it makes is that the ability to form bindings between partners and conversational material should be an attentionally limited process. Thus, for example, in a conversation with many active conversational partners, if one person produces a new label like "the dancer" to refer to a tangram figure, attentional constraints would predict that only a subset of these partners would be bound to the object label, perhaps depending on whom the speaker was attending to at the time of encoding (see Gleitman, January, Nappa, & Trueswell, 2007 for a related finding). Evaluating this hypothesis may benefit from examination of attentional distribution during unscripted conversation (Richardson et al., 2007), and relating these attentional processes to the subsequent ability to access partner-specific common ground.

3.3 Conversational Partners as a Meaningful Contextual Cue

At the beginning of this chapter we argued that people serve as a special type of context in language processing, such that conversational partners provide a contextual memory cue for the retrieval of jointly experienced information (Godden & Baddeley, 1975; Horton & Gerrig, 2005a; Smith & Vela, 2001). However, unlike the contextual cues typically studied in memory research, such as a room or a scent, conversational partners have a nonarbitrary, structured, and meaningful relationship to what is said in conversation. Conversational partners as individuals are also predictive of the form and contents of future conversational language. As a result, we argue that this meaningful and predictive nature of the partner should result in preferential encoding of partner-specific contextual information during conversational language use.

Unlike the arbitrary associations that one might form between lexical items and visual stimuli that are repeatedly paired (e.g., Marcus, Fernandes, & Johnson, 2012; Smith et al., 2014; Werker, Cohen, Lloyd, Casasola, & Stager, 1998), individuals can also become associated with objects in the world in a more meaningful way. When inanimate referents (e.g., a peanut) are established in a discourse context as having animate properties (e.g., a peanut in love), they take on such properties in subsequent processing of the language (Nieuwland & van Berkum, 2006), suggesting that adults have rich semantic representations about what people are like and can flexibly deploy them to understand language in novel ways. For instance, when one sees a man holding a bouquet, one is likely to encode not just the physical copresence of a man and some flowers, but also that he intends to give those flowers to a significant other, perhaps because of a special occasion. While it is debated whether infants and young children integrate this intentionality into, e.g., the word learning process (cf. Samuelson & Smith, 1998; Diesendruck, Markson, & Akhtar, 2004), in adulthood, it is likely that the encoding of a situation such as the one described above would go beyond simple co-occurrence of, e.g., man and flowers.

Motivation to encode the conversational partner as part of the relevant context may be due to a reasonable expectation that a given individual might say the same thing again, whereas other types of contexts (e.g., the weather) may be more loosely predictive of the contents of talk. As a result, encoding such intentional relationships could be adaptive in that they could facilitate prediction and processing in subsequent language understanding (see Borovsky & Creel, 2014; Creel, 2014; Ferguson & Breheny, 2011). Indeed, speakers tend to be consistent in their talk, being more likely to revisit a topic introduced by themselves than by another person in a conversation (Knutsen & LeBigot, 2012). Speakers are also consistent in how they describe objects over time, even when doing so is at the expense of adapting to the current context (Brennan & Clark, 1996; Heller & Chambers, 2014; Van der Wege, 2009). Moreover, listeners expect speakers to be consistent (Metzing & Brennan, 2003; Brown-Schmidt, 2009a; Brennan & Hanna, 2009; cf. Shintel & Keysar, 2007).

Support for the idea that conversational partners provide a memorial benefit by virtue of their meaningfulness comes from work in the memory literature, where a change in a meaningful context (e.g., removal of a context word semantically related to the target) is particularly detrimental to word recall (Pan, 1926). Interestingly, however, people as contexts provide a potent memory cue, even in typical memory tasks where interpersonal

communication is not particularly task-relevant. In a meta-analysis of environmental context effects in memory retrieval (i.e., the drop in memory performance when study and test are at different locations), whether or not the experimenter was the same person at learning and test significantly modulated the size of the environmental context effect (same experimenter: $d = 0.26$; different: $d = 0.62$; Smith & Vela, 2001). The fact that memory retrieval is impaired by the presence of a different experimenter suggests that even in cases where the identity of the experimenter is irrelevant to the studied material, it may, nonetheless, have an impact on the formation of memory representations (also see cf. Brown-Schmidt & Horton, 2014; Horton, 2007). An open question, then, is whether the influence of the partner would be stronger when that person is relevant to the studied material. According to our limited representations view, the conversational partner (or experimenter, in the case of memory research) is more likely to be encoded along with the relevant information when the partner is communicatively relevant.

If being meaningfully related to the language does, in fact, provide an additional contribution to the formation of partner-specific representations, we might expect the effect of the conversational partner to be modulated by his or her role in the interaction. Consider the simple conversational situation that arises when ordering a cup of coffee at a café. In this context, the cashier as an individual is somewhat immaterial to the exchange, as in this routinized experience the cashier is playing a role that could be satisfactorily fulfilled by many different individuals. The words exchanged with the café cashier are approximately the same whether it is always the same person behind the register or various people fill that role at different times. If the degree to which partners serve as a contextual cue depends on their relevance in the situation, we would expect partner-specific effects to be somewhat attenuated in such situations. Thus, similar to findings that speakers are more sensitive to the addressee perspective when instructing rather than informing (Yoon et al., 2012), the situations in which persons become strongly bound to discourse content may be those in which the other person is an important determinant of the way language is used. Testing this prediction would likely require examining the influence of communicative relevance on partner specificity in language use. If communicative relevance is a constraint on learning partner-specific bindings, we would expect better learning of partner-specific information, such as which individual knows the tangram label "the dancer," in situations where the conversational partner is relevant to the talk at hand (as in when jointly choosing a movie to watch) vs situations where the partner is more interchangeable (as in ordering coffee).

3.3.1 When Conversational Partners are Less Relevant

The idea that conversational partners are more strongly bound in memory to linguistic representations when communicatively relevant is consistent with a pattern of data in the literature which shows that partner-specific effects are attenuated or absent in situations with limited partner interaction (Barr, 2008; Barr & Keysar, 2002; Brown & Dell, 1987; Brown-Schmidt, 2009a; Kronmüller & Barr, 2007). By contrast, paradigms which encourage extensive, live interaction between the participants tend to show large effects of partner sensitivity (Lockridge & Brennan, 2002; Brown-Schmidt, 2009a; Hanna et al., 2003; Heller et al., 2008) that increase as speakers learn how to adjust to their partner throughout the conversation (Horton & Gerrig, 2002). In one of the few papers to explicitly test whether interactivity matters in the domain of partner-specific effects, Brown-Schmidt (2009a) examined the influence of interactivity per se on the processing of collaboratively established terms such as "the dancer" in situations such as the one depicted in the left panel of Figure 2.

Previous research had demonstrated that once a collaborative term has been established between two partners, in subsequent situations, if the original partner uses a new term to refer to the same image, e.g., "the falling-down man" to refer to the same object, interpretation is significantly delayed (Metzing & Brennan, 2003). However, the same penalty is not observed when a new conversational partner uses the new term, an effect which is interpreted as the participant addressee not extending an expectation for established terms to individuals who could not possibly be aware of them. Brown-Schmidt (2009a) examined a similar paradigm but contrasted the processing of these terms in live conversation, vs situations in which experimental participants listened to recordings. A critical design element was that the recordings were taken from previous recordings of utterances like "the dancer" and "the falling-down man" from other participants in the live conversation condition. Thus the critical auditory stimuli were the same in the live and prerecorded conditions, except that it was only the live conversation participants that had a chance to collaboratively and interactively establish these names. The results of these experiments showed that the partner-specific processing of old labels, e.g., "the dancer," was observed in live conversation, but not when participants listened to recordings. A small partner-specific effect for new labels, e.g., "the falling-down man," was also only observed in live conversation. When participants listened to recordings, they did not adjust processing of these labels in a partner-specific manner. These findings fall quite naturally out of a view of

partner-specific effects in conversation in which only those contextual cues that are communicatively relevant are encoded. On that view, the link in memory between the original partner, and the object label "the dancer" was only formed in the interactive conversation. By contrast, for participants in the noninteractive settings, the partner was not salient in the conversational situation because participants could not interact with him or her. As a result the partner was considered largely irrelevant and therefore partner-specific bindings were not learned.

A related set of findings comes from studies of multiparty conversation, in which the speaker may choose to design an utterance for one of several potential addressees. Thus far we have argued that encoding of partner-specific contextual cues is limited, and more likely when the partner is communicatively relevant. These attentional limits explain, for example, why an overhearer is not strongly bound in memory to the contents of the conversation. Likewise, if a former addressee were present as an overhearer in a subsequent situation, if communicative relevance is critical, information associated with that overhearer should not influence a speaker's choices. By contrast, if associations with all individuals in a conversation were retrieved, regardless of whether an individual was the intended addressee, this would predict that speakers should have difficulty designing utterances for one person out of a group of potential addressees when the common ground shared between the speaker and each of those individuals differs. Yoon and Brown-Schmidt (2014b) examined situations exactly such as this one, but found that referential patterns were guided entirely by the knowledge state of the intended addressee; there was no evidence that overhearers affected referring.

Likewise, in situations with two addressees, one of whom is knowledgeable, and the other of whom is unknowledgeable, Yoon and Brown-Schmidt (2014b) found that speakers designed expressions in a similar way as when addressing the unknowledgeable addressee alone, suggesting that the presence of the knowledgeable person did not automatically activate common ground. Taken together, these findings show that communicative relevance plays a role in both the formation and retrieval of partner-specific memory bindings. The presence of an individual in a communicative setting does not seem to result in automatic encoding or subsequent retrieval of associations in memory between that person and the information at hand.

3.4 Interim Summary

In summary, we propose that the number of links in memory between studied information and context is limited by attention (Simons & Chabris,

1999; Saiki, 2003; Franconeri et al., 2012), and memory (Eich, 1985; Smith, 1979), and by the communicative relevance of these contextual cues. The formation of only a limited number of associations can explain why partner effects are weak in noninteractive paradigms (Brown & Dell, 1987; Brown-Schmidt, 2009a; Barr, 2008; Barr & Keysar, 2002; Kronmüller & Barr, 2007; cf. Horton, 2007; Experiment 1; Brown-Schmidt & Horton, 2014), and why common ground is not formed with overhearers (Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992). This limited representations view is also consistent with findings that speakers are not affected by the presence of a nonaddressee with different common ground (Yoon & Brown-Schmidt, 2014b). In suggesting conversationally motivated limits on the formation of partner-as-context memories, this limited representations proposal bears similarity to the core of Clark and Marshall's (1978) idea, in that common ground is inferred only when the copresence of partner and the relevant information in the context is jointly attended.

4. LOOSE ENDS AND FUTURE QUESTIONS

4.1 Partner-Specific Bindings: How are They Learned?

Given that partner-specific representations are formed in conversation, a relevant question becomes what learning mechanisms govern the acquisition of these partner-specific bindings. One possibility is that binding a referential label such as "the dancer" to a specific partner is based on activation of associations between partners and labels (e.g., Horton & Gerrig, 2005a; Horton, 2007), which Pickering and Garrod (2004) propose allows for the automatic formation of implicit common ground. Alternatively, learning of such partner-specific bindings may be supported by error-based learning (Chang, Dell, Bock, & Griffin, 2000; also see Fine & Jaeger, 2013 for an example distinction between activation vs error-based learning accounts of syntactic priming).

One way of teasing apart an association-based learning view from an error-based learning view would be to examine the magnitude of learning following discourse turns that were more vs less surprising (for a discussion of surprisal in sentence processing see Jaeger & Snider, 2013; Fine, Jaeger, Farmer, & Qian, 2013; Levy, 2008; Hale, 2003). When a familiar partner uses a novel term for a game piece, such as "the falling-down man," when "the dancer" was expected, this should be particularly surprising to the addressee; indeed, such expressions are typically interpreted with difficulty (Metzing & Brennan, 2003). If so, on an error-based learning account

this should result in particularly strong learning of the term "the falling-down man" compared to a situation in which the same expression is used by a new partner, and therefore less surprising. By contrast, on an association-based account, the amount of association between the label and the partner should increase for both old and new partners to a similar degree. Examining the amount of change in the listener's expectation for the new term before vs after hearing the unexpected expression should provide a means of distinguishing between these two hypotheses. On an association-based learning account, the amount of change should be equivalent regardless of how surprising the new expression was. By contrast, on an error-based learning account, the predicted amount of change in expectation should increase with greater surprisal.

A key goal for future research in perspective-taking is the development of implemented models of this process; such models would dramatically improve the precision and testability of current theories (e.g., see Heller, Parisien, & Stevenson, 2012). Understanding the relevant learning mechanisms involved in encoding of partner-specific bindings is an important first step toward the development of such models.

4.2 Domains of Partner-Specific Knowledge

Much of the present discussion has focused on representations of the beliefs and knowledge of conversational partners, and how this information is learned and used in the service of language processing. Experimental research, however, shows that conversational participants are sensitive to a wide variety of partner-specific features and attributes. Thus an open question is whether each of these types of knowledge are learned and used in similar ways in conversation, or whether they differ fundamentally. For example, is knowledge of the acoustic properties of an individual talker's speech such as their accent (Dahan et al., 2008) encoded and used through similar mechanisms as knowledge of what a talker can and cannot see (Horton & Keysar, 1996)? In other words, is knowing that a speaker knows something, such as the name of an object, different than associating them with something, such as an accent? The speaker-knowledge case differs from the speaker-accent case in that, whereas both involve linking an individual to a particular piece of information, in the knowledge case we extend metacognitive states to that person as well. That is, not only is the speaker associated with a given piece of information but you know they know that information as well.

Whether these types of information function similarly or differently in language use may depend on the domain of inquiry. For example, in a situation where the talker's knowledge state is less relevant to the communicative context, such as in a situation where the talker is reading from a book, knowledge of a talker's accent may guide lexical access more than what a talker does and does not know (as, after all, when reading from a book, the content of what is said is not limited to what the talker already knows). By contrast, in a domain more tightly linked to talker knowledge, as in a situation where the talker is asking the addressee to hand her items in order to bake a cake (Hanna & Tanenhaus, 2004), which ingredients the talker is and is not aware of may play a larger role.

Indeed, robust effects of talker accent are observed even in cases where participants are listening to a recording (Trude & Brown-Schmidt, 2012; Dahan et al., 2008), whereas effects of what a talker does and does not know are attenuated with recordings (Brown-Schmidt, 2009a; Lockridge & Brennan, 2002). An open question, then, is whether talker–accent (Dahan et al., 2008) and talker–word (Creel et al., 2008) pairings might similarly have more of an effect in live conversation where a co-present talker may provide a more salient memory cue for learning such pairings. If not, it may point to these types of talker-based pairings resulting from a different type of learning process.

Related to these questions is how these different domains of talker knowledge are structured in memory. Some findings suggest that when talker–object associations are distinctive, speakers have more success in using them (Horton & Gerrig, 2005b). However, others find little improvement in audience design when partners are associated with structurally distinct information (Heller et al., 2012; Gorman et al., 2013). If this knowledge is structured, an open question is whether conversational partners make use of correlated cues. For example, if accent is predictive of grammatical preferences, does exposure to a new talker's accent result in a change in syntactic predictions? Answering these questions about the domains of partner knowledge, and the structure of this knowledge in memory, will likely provide explanatory power in understanding when partner-specific effects are and are not likely to emerge in language use.

4.3 Participant Role in Conversation

Well established in the memory literature are findings that produced information is remembered better than heard information. The production effect (MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010; MacLeod, 2011;

Forrin, MacLeod, & Ozubko, 2012; Knutsen & LeBigot, 2012) refers to the phenomenon that saying, mouthing, spelling, writing, and typing out information all result in better memory performance than receptive encoding, such as reading silently. By contrast, in conversational settings, some evidence suggests better recall of what is heard than what is told (Stafford & Daly, 1984), possibly due to a response bias to report egocentrically new information (that which was told to you). Differential situational relevance of spoken vs heard information may play a role as well; partners tend to overestimate their own contributions to the conversation when those contributions were helpful to solving the problem at hand (Ross & Sicoly, 1979). These findings suggest that speakers and addressees are unlikely to have parallel or aligned representations of the conversational history (cf. Pickering & Garrod, 2004). Instead, individuals are likely to better encode what they themselves said in conversation, and to exhibit biases in recall based on the relevance of that information.

Related to these findings of differential recall of conversational content ("item" memory), conversational role may additionally influence the binding of item to information source (the speaker) and recipient or "destination" (the addressee). Gopie and MacLeod (2009) compared memory for items, sources, recipients, and item–source and item–recipient bindings in a paradigm in which 60 facts were paired with 60 pictures of faces. In the destination memory condition, participants silently read a fact, then saw a face, and were instructed to tell the fact to the face. In the source memory condition, participants saw a face, and then read a printed fact with the instruction to imagine the person pictured was telling them the fact. Memory for faces and for face–fact pairings was higher in the source memory condition (memory for who told you a specific fact) compared to the destination memory condition (memory for who you told a specific fact to), whereas memory for the facts was equivalent (cf. Brown, Jones, & Davis, 1995). Thus unlike the production effect, in which item-level memory is better for what is spoken, these findings, somewhat paradoxically, show that binding of person to item is better for what is heard.

While this experimental situation is rather divorced from typical conversational settings, if the results were to generalize to typical conversation, it would suggest that speakers and listeners may come away from the conversation with different representations of the discourse history (Yoon & Brown-Schmidt, 2013). Such a finding would carry theoretical significance as it would indicate that basic memory processes constrain the degree to which interlocutors can develop coordinated representations of the

discourse history. Given that the shared discourse history is one of the key contributors to the formation of common ground (Clark & Marshall, 1978), this would mean that there are fundamental limits on the degree to which conversational partners can have common ground. Understanding the impact of such limits on conversational language use remains a key goal for future research. Lastly, as we consider how memory for conversation depends on role, similar questions emerge in the domain of multiparty conversation where different individuals may be active at different points in the conversation, and different pairs of individuals may be more relevant at some points than others. If the binding of person to item in conversation is, in fact, better for addressees (Gopie & MacLeod, 2009), this might suggest that addressees would be in a better position to monitor who knows what in a multiparty conversation, compared to speakers. If so, this should result in asymmetries in information management across the course of a dialog. These may be revealed through systematic monitoring of audience design in unscripted conversation as a function of participant role when the relevant information was introduced into the discourse.

5. CONCLUSIONS

Language use in conversation is guided not only by local context but also by representations of the conversational partner's perspective on that context. These perspective representations are stored in memory and are used to guide subsequent language use in a partner-specific manner. Thus, in much the same way that studied information is bound in memory to the environmental context of learning, conversational partners are bound to the information discussed in conversation. These partner-specific contextual bindings support basic conversational processes such as the act of describing an object in a way that one's addressee can easily understand, based on the shared conversational history.

Extensive research has demonstrated successful use of a wide range of partner-specific contextual information in conversation, such as learning and using detailed information about what a specific partner can see, what they know from past experience, as well as their spatial perspective, preferences, accent, prosody, and goals. Taken together, these findings show remarkable sensitivity to an enormous range of potential bindings between person or context, and linguistic information. Yet, if in every conversational situation, conversational partners were to encode all these potential contextual bindings, the number of encoded bindings would be unbounded, and beyond

the limits of attention and memory. Here we propose a limited representations view of the encoding of partner-specific bindings in memory in which the number of bindings between contextual elements (including the partner) and the language at hand is limited by attentional and memorial constraints.

The idea that the encoding of contextual bindings during conversation is limited provides a cognitively plausible explanation of why conversational partners fail to guide language use in a partner-specific manner in noninteractive settings, and why speakers are able to tailor language to the intended addressee, even in situations where overhearers or other nonaddressees are present in the context. In brief, the predictions of this view are that partners should be more tightly bound to context during encoding when the partner is communicatively relevant, or meaningfully related to the information under discussion. In multiparty conversation, bindings of partner to context may be limited to a subset of all possible addressees, in particular those addressees whom the speaker was attending to at the time of production, and vice versa. Finally, encoding and retrieval of partner-specific contextual bindings should be under the metacognitive control of conversational participants such that intentional integration of partner with information during encoding and/or retrieval should produce stronger partner-specific effects.

In conclusion, successful communication in conversation relies on the integration of both contextual and perspective representations into the processes of language production and comprehension. The extent to which partner-specific representations of perspective are encoded, however, is limited by basic cognitive processes of attention and memory. Understanding the scope and structure of representations that are formed clarifies potentially conflicting findings in the literature on perspective-taking, and offers insights into when perspective-taking will and will not be successful.

REFERENCES

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science, 262*, 2042–2044.

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say *thee uh* you are describing something hard: the on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 914–930. http://dx.doi.org/10.1037/0278-7393.33.5.914.

Arnold, J. E., Tanenhaus, M. K., Altmann, R., & Fagnano, M. (2004). The old and thee, uh, new. *Psychological Science, 15*, 578–582. http://dx.doi.org/10.1111/j.0956-7976.2004.00723.x.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science, 27*, 195–225.

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of memory and language, 59*, 457–474.
Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391–418.
Barr, D., & Seyfeddinipur, M. (2010). The role of fillers in listener attributions for speaker disfluency. *Language and Cognitive Processes, 25*, 441–455. http://dx.doi.org/10.1080/01690960903047122.
Borovsky, A., & Creel, S. C. (2014). Children and adults integrate talker and verb information in online processing. *Developmental Psychology, 50*, 1600–1613.
Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics, 61*, 206–219.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 482–493.
Brennan, S. E., & Hanna, J. E. (2009). Partner-Specific adaptation in dialog. *Topics in Cognitive Science, 1*, 274–291.
Bromme, R., Jucks, R., & Wagner, T. (2005). How to refer to 'Diabetes'? Language in online health advice. *Applied Cognitive Psychology, 19*, 569–586.
Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: the explicit mention of instruments. *Cognitive Psychology, 19*, 441–472.
Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational source monitoring. *Psychology and Aging, 10*, 111.
Brown-Schmidt, S. (2009a). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*, 171–190.
Brown-Schmidt, S. (2009b). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*, 893–900.
Brown-Schmidt, S. (2012). Beyond common and privileged: gradient representations of common ground in real-time language use. *Language and Cognitive Processes, 27*, 62–89.
Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition, 107*, 1122–1134.
Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: incremental perspective-taking in language processing. *Dialog and Discourse, 2*, 11–33.
Brown-Schmidt, S., & Konopka, A. E. (2011). Experimental Approaches to referential domains and the on-line processing of referring expressions in unscripted conversation. *Information, 2*, 302–326.
Brown-Schmidt, S., & Tanenhaus, M. K. (2006). Watching the eyes when talking about size: an investigation of message formulation and utterance planning. *Journal of Memory and Language, 54*, 592–609.
Brown-Schmidt, S., & Horton, W. S. (2014). The influence of partner-specific memory associations on picture naming: a failure to replicate horton (2007). *PloS One, 9*(10), e109035. http://dx.doi.org/10.1371/journal.pone.0109035.
Buxó-Lugo, A., Toscano, J., & Watson, D. G. (March 2013). Task effects on prosodic prominence. In *Paper presented at the 26th Annual CUNY conference on Human sentence processing, Columbia, SC*.
Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Carlson, G. N., & Filip, H. (2002). Circumscribing referential domains during real time language comprehension. *Journal of Memory and Language, 47*, 30–49.
Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: a comparison of models of sentence production. *Journal of Psycholinguistic Research, 29*, 217–229.

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.
Clark, H. H. (1996). *Using language*. Cambridge University Press.
Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on socially shared cognition, 13*, 127–149.
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. (1978). Reference diaries. In D. L. Waltz (Ed.), *TINLAP-2: Theoretical Issues in Natural language Processing-2. Association for computing Machinery, New York* (pp. 57–63).
Clark, H. H., & Marshall, C. M. (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: Cambridge University Press.
Clark, H. H., & Wasow, T. (1988). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201–242.
Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.
Creel, S. C. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development, 83*, 2042–2056.
Creel, S. C. (2014). Preschoolers' flexible use of talker information during word learning. *Journal of Memory and Language, 73*, 81–98.
Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition, 106*, 633–664.
Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2012). Word learning under adverse listening conditions: context-specific recognition. *Language and Cognitive Processes, 27*, 1021–1038.
Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*, 264–285.
Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition, 108*, 710–718.
Diesendruck, G., Markson, L., Akhtar, N., & Reudor, A. (2004). Two-year-olds' sensitivity to speakers' intent: an alternative account of Samuelson and Smith. *Developmental Science, 7*, 33–41.
Doumas, L. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological review, 115*, 1–43.
Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 764–770.
Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54*, 554–573.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.
Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behavior based on complex, conflicting desires. *Cognition, 119*, 179–196.
Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48*, 257–263.
Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30*, 210–233.
Fine, A. B., & Florian Jaeger, T. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37*, 578–591.
Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one, 8*, e77661.
Fletcher, C. R., Hummel, J. E., & Marsolek, C. J. (1990). Causality and the allocation of attention during comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 233–240.

Forrin, N. D., MacLeod, C. M., & Ozubko, J. D. (2012). Widening the boundaries of the production effect. *Memory & Cognition, 40*, 1046–1055.
Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition, 122*, 210–227.
Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: for the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.
Galati, A., Michael, C., Mello, C., Greenauer, N. M., & Avraamides, M. N. (2013). The conversational partner's perspective affects spatial memory and descriptions. *Journal of Memory and Language, 68*, 140–159.
Gann, T. M., & Barr, D. (2012). Speaking from experience: audience design as expert performance. *Language and Cognitive Processes, 29*, 744–760. http://dx.doi.org/10.1080/01690965.2011.641388.
Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.
Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the *give* and *take* between event apprehension and utterance formulation. *Journal of memory and language, 57*, 544–569.
Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology, 66*, 325–331.
Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of Psychology, 71*, 99–104.
Goldinger, S. D. (1998). Echoes of echoes? an episodic theory of lexical access. *Psychological review, 105*, 251–279.
Gopie, N., & MacLeod, C. M. (2009). Destination memory: stop me if I've told you this before. *Psychological Science, 20*, 1492–1499.
Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: speakers' choice of referring expressions reflects shared experience. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*, 843–853.
Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology. General, 121*, 480–506.
Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics, 12*, 175–204.
Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32*, 101–123.
Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: evidence from eye movements. *Cognitive Science, 28*, 105–115.
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.
Heller, D., & Chambers, C. G. (2014). Would a *blue kite* by any other name be just as blue? Effects of descriptive choices on subsequent referential behavior. *Journal of Memory and Language, 70*, 53–67. http://dx.doi.org/10.1016/j.jml.2013.09.008.
Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: shared knowledge affects referential form. *Topics in Cognitive Science, 4*, 290–305.
Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*, 831–836.
Heller, D., Parisien, C., & Stevenson, S. (2012). Perspective-taking behavior as the probabilistic weighing of multiple domains. In *Poster presented at the city University of New York conference on human sentence processing, New York, NY*.
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. (2002). Speakers' experiences and audience design: knowing when and how to adjust utterances to addressees. *Journal of Memory and Language, 47*, 589–606.
Horton, W. S., & Gerrig, R. J. (2005a). The impact of memory demands on audience design during language production. *Cognition, 96*, 127–142.
Horton, W. S., & Gerrig, R. J. (2005b). Conversational common ground and memory processes in language production. *Discourse Processes, 40*, 1–35.
Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.
Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: the influence of speaker-specific memory associations on reference resolution. *Memory & Cognition, 40*, 113–126.
Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*, 281–290.
Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9*, 90–98.
Isaacs, E., & Clark, H. H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*, 26–37.
Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*, 57–83.
Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological bulletin, 114*, 3.
Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition, 124*, 66–71.
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*, 32–37.
Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*, 25–41.
Knutsen, D., & Le Bigot, L. (2012). Managing dialogue: how information availability affects collaborative reference production. *Journal of Memory and Language, 67*, 326–341.
Konopka, A. E., & Brown-Schmidt, S. (2014). Message encoding. In V. Ferreira, M. Goldrick, & M. Miozzo (Eds.), *The Oxford handbook of language production*.
Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343–346.
Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*, 436–455.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126–1177.
Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*, 550–557.
MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9*, 157–201.
MacLeod, C. M. (2011). I said, you said: the production effect gets personal. *Psychonomic Bulletin & Review, 18*, 1197–1202.
MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 671–685. http://dx.doi.org/10.1037/a0018785.
Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human perception and performance, 33*, 391–409.
Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2012). The role of association in early word-learning. *Frontiers in Developmental Psychology, 3*, 1–6.

Massaro, D. W., & Anderson, N. H. (1971). Judgmental model of the Ebbinghaus illusion. *Journal of Experimental Psychology, 89*, 147–151.
Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2006). The effect of perceptual availability and prior discourse on young children's use of referring expressions. *Applied Psycholinguistics, 27*, 403–422.
Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and Decision making. *Annual Review of Psychology, 49*, 447–477.
Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*, 201–213.
Mulligan, N. W. (2011). Conceptual implicit memory and environmental context. *Consciousness and cognition, 20*, 737–744.
Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*, 329–336.
Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of cognitive neuroscience, 18*, 1098–1111.
Olson, D. R. (1970). Language and thought: aspects of a cognitive theory of semantics. *Psychological review, 77*, 257–273.
Osgood, C. E. (1971). Exploration in semantic Space: a personal diary. *Journal of Social Issues, 27*, 5–64.
Pan, S. (1926). The influence of contextual conditions upon learning and recall. *Journal of Experimental Psychology, 9*, 468–491.
Parker, A., Dagnall, N., & Coyle, A. M. (2007). Environmental context effects in conceptual explicit and implicit memory. *Memory, 15*, 423–434.
Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics, 27*, 89–110.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences, 27*, 169–190.
Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science, 18*, 407–413.
Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33*, 1468–1482.
Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37*, 322–336.
Roth, J. C., & Franconeri, S. L. (2012). Asymmetric coding of categorical spatial relations in both language and vision. *Frontiers in Psychology, 3*, 1–14.
Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L., & Nguyen, E. (2014). Visuospatial perspective-taking in conversation and the role of Bilingual experience. *Journal of Memory and Language, 74*, 46–76.
Saiki. (2003). Feature binding in object-file representations of multiple moving items. *Journal of Vision, 3*, 6–21.
Samuelson, L. K., & Smith, L. B. (1998). Memory and attention make smart word learning: an alternative account of Akhtar, Carpenter, and Tomasello. *Child Development, 69*, 94–104.
Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1–24.
Schober, M. F. (2009). Spatial dialogue between partners with mismatched abilities. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds.), *Spatial language and dialogue* (pp. 23–39). Oxford: Oxford University Press.
Schober, M. F., & Brennan, S. E. (2003). Processes of interactive spoken discourse: the role of the partner. *Handbook of discourse processes*, 123–164.
Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*, 211–232.

Searle, J. (1969). *Speech acts*. Cambridge University Press.
Sedivy, J. C. (2005). Evaluating explanations for referential context effects: evidence for Gricean mechanisms in online language interpretation. 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* (pp. 153–171). Cambridge, MA: MIT press.
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, and Cognition, 33*(2), 357–369. http://dx.doi.org/10.1037/0278-7393.33.2.357.
Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception-London, 28*, 1059–1074.
Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 460–471.
Smith, S. M. (1982). Enhancement of recall using multiple environmental contexts during learning. *Memory & Cognition, 10*, 405–412.
Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342–353.
Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in cognitive sciences, 18*, 251–258.
Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic bulletin & review, 8*, 203–220.
Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication Research, 10*, 379–402.
Stalnaker, R. C. (1978). Assertion. *Syntax and Semantics, 9*, 315–332.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. E. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 632–634.
Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller, & P. Eimas (Eds.), *Handbook of cognition and perception*. San Diego, CA: Academic Press.
Trude, A. M., & Brown-Schmidt, S. (2012). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes, 27*, 979–1001.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience, 20*, 580–591.
Van der Wege, M. M. (2009). Lexical entrainment and lexical differentiation in reference phrase choice. *Journal of Memory and Language, 60*, 448–463.
Walker, M. A. (1996). Limited attention and discourse structure. *Computational Linguistics, 22*, 255–264.
Wardlow-Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speaker's control over leaking private information during language production. *Psychological Science, 17*, 273–277.
Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14- month-old infants. *Developmental Psychology, 34*, 1289–1309.
Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31*, 183–194.
Yee, E., & Heller, D. (2012). Looking more when you know less: goal-dependent eye movements during reference resolution. In *Poster presented at the Annual Meeting of the Psychonomic Society, Minneapolis, MN*.
Yeh, W., & Barsalou, L. W. (2006). The situated nature of concepts. *American Journal of Psychology, 119*, 349–384.

Yoon, S. O., & Brown-Schmidt, S. (2013). Lexical differentiation in language production and comprehension. *Journal of Memory and Language, 69*, 397–416.

Yoon, S., & Brown-Schmidt, S. (2014a). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 919–937.

Yoon, S. O., & Brown-Schmidt, S. (2014b). Aim low: speakers design utterances for the most naïve addressee. In *Poster presented at CUNY 2014, Columbus, OH*.

Yoon, S. O., Koh, S., & Brown-Schmidt. (2012). Influence of perspective and goals on reference production in conversation. *Psychonomic Bulletin & Review, 19*, 699–707.