Mem Cogn (2017) 45:1281–1294
DOI 10.3758/s13421-017-0730-3

Memory for conversation and the development of common ground

Geoffrey L. McKinley[1] · Sarah Brown-Schmidt[1,2,3] · Aaron S. Benjamin[1,2]

Published online: 6 July 2017

Abstract Efficient conversation is guided by the mutual knowledge, or common ground, that interlocutors form as a conversation progresses. Characterized from the perspective of commonly used measures of memory, efficient conversation should be closely associated with item memory—what was said—and context memory—who said what to whom. However, few studies have explicitly probed memory to evaluate what type of information is maintained following a communicative exchange. The current study examined how item and context memory relate to the development of common ground over the course of a conversation, and how these forms of memory vary as a function of one's role in a conversation as speaker or listener. The process of developing common ground was positively related to both item and context memory. In addition, content that was spoken was remembered better than content that was heard. Our findings illustrate how memory assessments can complement language measures by revealing the impact that basic conversational processes have on memory for what has been discussed. By taking this approach, we show that not only does the process of forming common ground facilitate communication in the present, but it also promotes an enduring record of that event, facilitating conversation into the future.

✉ Geoffrey L. McKinley
mckinle2@illinois.edu

[1] Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel St., Champaign, IL 61820, USA

[2] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Champaign, USA

[3] Department of Psychology and Human Development, Vanderbilt University, 230 Appleton Place # 552, Nashville, TN 37203, USA

Keywords Common ground · Destination memory · Source memory · Egocentrism · Referential communication · Conversation

In conversation, knowledge of what information is and is not mutually known is central to having a successful conversation (Clark, 1996). Throughout the course of a conversation, the amount of mutual knowledge, or common ground, grows as conversational partners exchange new information. Common ground promotes efficient communication, as it allows the speaker to take advantage of the addressee's knowledge in constructing effective and non-redundant utterances (Clark & Schaefer, 1989; Wilkes-Gibbs & Clark, 1992). A key goal at the interface of memory and language research is understanding how this process of forming common ground influences later memory for conversation. After all, successful communication often involves informational exchange in the moment as well as memory for that information later on.

Evidence for the growth of common ground comes from classic studies in which conversational partners repeatedly refer to a set of hard-to-name referents. A now well-replicated finding is that when these referents are first mentioned, they are referred to with descriptions of considerable length in order to minimize ambiguity. However, given multiple opportunities to refer to each item, the partners develop brief, unique labels for each one (Clark & Wilkes-Gibbs, 1986; Krauss, Garlock, Bricker, & McMahon, 1977; Krauss & Weinheimer, 1966; Wilkes-Gibbs & Clark, 1992), a process known as lexical entrainment (Brennan & Clark, 1996). When a speaker subsequently addresses a different, naïve partner, they typically use distinct, longer expressions (Wilkes-Gibbs & Clark, 1992; also see Brennan & Clark, 1996). This latter finding shows that speakers are sensitive to the knowledge state of the addressee; that is, speakers engage in audience design (Schober & Clark, 1989). Much of the work on common ground comes from studies that use referential communication tasks (Krauss & Weinheimer, 1966). In such tasks, a speaker must describe a set of pictures to an addressee, and the addressee's task is to rearrange the same set of pictures on his or her screen to match the arrangement of the speaker's. Importantly, the speaker must be able to describe each picture with a sufficient amount of detail so that the addressee can identify the correct referent. In response, the addressee may give some indication of whether or not the description was adequate, and perhaps participate in forming a description. It should be emphasized that even in cases where the addressee says few words, those words serve an important function of grounding that the speaker's utterance was understood (Clark & Brennan, 1991).

Although the central role of memory in developing common ground is widely recognized (e.g., Barr & Keysar, 2002; Brown-Schmidt, 2012; Clark & Marshall, 1981; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Haviland & Clark, 1974; Horton, 2007; Horton & Gerrig, 2005a, 2005b), little research has explicitly evaluated memory for events within conversation (e.g., Fischer, Schult, & Steffens, 2015; Fraundorf, Watson, & Benjamin, 2015; Keenan, MacWhinney, & Mayhew, 1977; Knutsen & Le Bigot, 2014; MacWhinney, Keenan, & Reinke, 1982; Stafford & Daly, 1984), and even less research has attempted to link those measures of memory to measures of the formation of common ground. The goal of the present study is to examine how memory is affected by the process of forming common ground and by one's role in that process. Specifically, we examine how the process of forming common ground impacts memory for items (the content of conversation) and memory for context (the sources and destinations of utterances within a conversation) and how this relationship varies as a function of one's conversational role as speaker or addressee.

According to one view, the efficiency with which common ground is accessed stems from associations that have been formed in memory between a particular discourse partner and the contents of talk (Horton & Gerrig, 2005a). Knowledge of having previously discussed a referent, for example, becomes accessible when the partner and referent serve as cues of sufficient specificity to effect accurate and efficient retrieval of this common ground. Indeed, Horton and Gerrig (2005a) found that distinctive memorial cues linking partners and referents result in more successful use of common ground. This result is consistent with the general view that source monitoring is enabled in part by environmental correlations between the nature of the information and the specific source (Johnson, Hashtroudi, & Lindsay, 1993).

Contributions of memory to conversation

Clearly, the task of describing items in ways that are specific to a given partner not only requires the speaker to remember the earlier utterance and its corresponding referent but also to remember the person that the utterance and referent are associated with. Remembering whom you discussed a particular referent with is a type of context memory. Consider a case in which the director in a referential communication task describes a picture to a matcher (e.g., "Click on the perched blue jay"). Using terminology from the memory literature, from the matcher's perspective, the director can be defined as the information source (e.g., Johnson et al., 1993). Conversely, from the director's perspective, the matcher can be defined as the information destination—that is, the person to whom an utterance was communicated to (Gopie & MacLeod, 2009). Here we examine a particular type of context memory in conversation—memory for the source and destination of utterances.

Extant research in the memory literature has asked the general question of whether or not destination memory differs from source memory, and the results are mixed. One finding that is relevant to this question is the generation effect. This effect describes the finding that generating an item out loud or reconstructing an item given partial information, as opposed to simply reading the item, will enhance memory for that item (Jacoby, 1978; Slamecka & Graf, 1978; see MacLeod, Gopie, Hourihan, Neary, & Ozubko, 2010, for a related effect). This effect is germane to conversation because one may remember an utterance better if self-generated, and so speakers should be more likely to remember what was said than listeners.

Whereas item memory—what was said—is likely to show a generation (speaker) benefit, it is not entirely clear whether there are benefits or costs of generation to remembering who was associated with an utterance. According to one perspective, generation diverts resources that would have otherwise been directed towards processing the context (Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; Koriat, Ben-Zur, & Druch, 1991). This view predicts that a speaker will have worse memory for the destination of their utterance than the listener will have for the source of the utterance. In contrast, according to a different view, variables such as generation may provide general benefits to memory, including contextual aspects of generated situations (E. J. Marsh, Edelman, & Bower, 2001). In conversation, this view would predict that speakers will have better utterance and context (partner) memory than listeners.

While some researchers have found better context memory following generation (Koriat et al., 1991, Experiment 1; E. J. Marsh et al., 2001), others have found worse memory for context following generation (Brown, Jones, & Davis, 1995; Fischer et al., 2015; Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; Koriat et al., 1991, Experiment 2). However, few of these studies were conducted in the context of live conversation (cf. Brown et al., 1995; Fischer et al., 2015; Jurica & Shimamura, 1999), making their relevance for conversation memory unclear. Given previous demonstrations of language processing differences between live conversation versus noninteractive language settings (Brown-Schmidt, 2009; Brown-Schmidt & Fraundorf, 2015), it is not obvious how these findings would generalize to a genuine conversational situation.

The reason it is difficult to generalize findings from simple memory tasks to conversation is that many of these studies control for characteristics that normally vary in live conversation. Conversation is a very dynamic situation that provides rich, multidimensional information. However, there are some studies in the language literature that have used both interactive designs similar to natural conversation and memory measures of the content of the conversation. For example, Stafford and Daly (1984) gave participants a free recall test following a spontaneous conversation and found that participants tended to recall more of their partner's contributions to a conversation than their own (see also Stafford, Burggraf, & Sharkey, 1987). This finding is interesting because of the contrast with the finding that participants are biased to reuse their own utterances within a conversation (Knutsen & Le Bigot, 2014), an egocentric preference that is thought to be a consequence of greater availability (Ross & Sicoly, 1979).

It should be noted that Stafford and Daly's (1984) finding is inconsistent with those reviewed earlier from the memory literature, further heightening concern over the role of task differences. For example, in Stafford and Daly's study, the topic of conversation was unconstrained, which relinquished experimental control over the nature of the information to be remembered. Perhaps, then, in unscripted conversation, what one's partner says is simply more memorable, leading to better recall of what is heard than what is said (Stafford & Daly, 1984). By contrast, in experiments that show a generation benefit, the materials are typically well controlled and comparable across the read and generate conditions (e.g., Slamecka & Graf, 1978). In one study that used a naturalistic storytelling procedure (Isaacs, 1990), a generation benefit was found— a finding that may owe to the structured nature of these stories. A further consideration is that Stafford and Daly's use of a free-report task, like recall, means that individuals had to make an explicit choice about what to report during the recall task. Participants may have remembered some aspects of the conversation that they chose to withhold on the test if they, for example, deemed it unimportant, irrelevant, or unflattering (Koriat & Goldsmith, 1996).

A standard way to deal with this problem of reporting bias is to use a memory measure that allows for the separation of the contributions of response bias and memory, like recognition memory (Banks, 1970; Egan, 1958; Green & Swets, 1966; Macmillan & Creelman, 2005). Recognition memory measures also provide experimental control over the order in which information from the conversation is tested. Consequently, this procedure mitigates concerns of list-strength effects (Ratcliff, Clark, & Shiffrin, 1990) and output interference (Roediger & Schmidt, 1980), both of which exaggerate the advantages of a more memorable class of items over a less memorable class.

Only two studies to date have simultaneously tested context memory with more than one external source (or destination) using a recognition-type test and an interactive conversational paradigm (Brown et al., 1995; Fischer et al., 2015). In Brown et al. (1995), participants were instructed to ask a question, respond to a question with an answer, or simply listen. A subsequent memory test probed for memory of what was discussed as well as context (partner) memory, and found that the responder's performance on identifying the questioner was worse than the questioner's identification of the responder. These findings suggest that context memory is compromised for the speaker. However, as E. J. Marsh et al. (2001) pointed out, the difficulty of the task was confounded with conversational role. That is, responder identification seemed to be easier than questioner identification, regardless of the role one occupied in the triad. What is needed is a procedure in which the contents of the conversation are controlled across role. Fischer et al. (2015) adapted this design using sentence fragments and found that memory was superior for sentence fragments that one had completed (a generation benefit) but that partner memory was worse for fragments that one had completed (a generation penalty). A limitation of this study, however, is that participants were not permitted to interact with each other beyond the sentence completions; it is unknown if this generation penalty for context memory would extend to more natural conversation.

Common ground and memory

The purpose of this work is to measure item and context memory in an interactive conversation paradigm, and then relate these measures to the process by which conversational partners form common ground. In referential communication tasks like the one used in the present research, the speaker initially provides long descriptions of each image. As the conversation progresses and common ground is formed, the conversational partners tend to settle on brief labels for each image. This process of accumulating common ground is known to facilitate future communication: Whereas addressees who have common ground for the image labels typically interpret them at near perfect accuracy, people who are new to the task, or who were not actively involved in forming common ground, perform more poorly (Schober & Clark, 1989). Indeed, theories of the role of memory for common ground rely heavily on the binding of discourse-relevant information with specific conversation partners (e.g., Horton & Gerrig, 2005a, 2005b). Although there is evidence that common ground supports communication (Richardson & Dale, 2005; Richardson, Dale, & Kirkham, 2007; Schober & Clark, 1989), the relationship between conversational memory and the development of common ground remains unexplored. We hypothesize that forming common ground should promote memory for what has been discussed, and with whom.

A related question is whether conversational partners form similar representations of the discourse history. If forming common ground promotes memory for speakers and addressees alike, then speakers and addressees should display comparable memory for past items and contexts in conversation. However, other work suggests that conversation is strongly influenced by each partner's own egocentric bias (Knutsen & Le Bigot, 2014); while this finding hints at a generation benefit for item memory in dialogue, Knutsen and Le Bigot (2014) found no evidence of a generation benefit in a free-recall measure of memory. Indeed, the conversational memory findings that use free recall as a response measure (e.g., Isaacs, 1990; Knutsen & Le Bigot, 2014; Miller, deWinstanley, & Carey, 1996; Ross & Sicoly, 1979; Stafford et al., 1987; Stafford & Daly, 1984) provide evidence that is mixed, in that some findings are inconsistent with the generation effect, which is known to be highly reliable (see MacLeod et al., 2010, for a discussion). If conversation is heavily driven by the egocentric biases of the interlocutors, then the act of generation may overwhelmingly improve memory, above and beyond any beneficial effect of forming common ground. Alternatively, the process of forming common ground may work to rectify asymmetries between speakers' and listeners' memory for conversation.

Experiment

The goal of the present research is to examine how the process of forming common ground influences memory for what has been discussed (item) and with whom (context). We used an established task—the referential communication task (Krauss & Weinheimer, 1966) that allows conversational partners to form common ground, followed by a recognition test that probes item memory and context memory. Participants were tested in groups of four at a time and were paired off into dyads to complete the referential communication task. Through repeated repairings, we created situations where each participant formed common ground with two other partners. We used a memory test that provides measures of item and context (partner) memory that are independent of response bias (Green & Swets, 1966; Macmillan & Creelman, 2005), and that can be related to the development of common ground for those items.

Method

Participants Eighteen groups, each composed of four individuals, participated in the study. Participants were recruited from introductory-level psychology classes from the University of Illinois at Urbana-Champaign and received partial course credit in exchange for their participation. Although participants were all fluent English speakers enrolled as students at a major university, no participant was excluded on the basis of their native language. As a result, the sample includes a diverse—and representative—sample of the undergraduate population that included nonnative speakers.[1]

The experiment was designed such that each group would be composed of four genuine participants. However, due to participant no-shows, this was not always feasible, and research assistants were needed to fill in. Out of these 18 groups, five contained one research assistant who took the place of one participant that did not show up. As a result, these groups were comprised of three genuine participants each. In two additional instances, only one genuine participant showed up to the experimental session, which precluded the need to have a group of four. In these instances, two research assistants participated, which resulted in three partners in total for these groups. Although the experiment was designed for four participants, the single participant's experience in these cases was the same as in the four-participant situation. In cases where research assistants filled in, we did not attempt to obscure their role as assistants, as we expected that their participation in their genuine role would lead to a more natural interaction (see Kuhlen & Brennan, 2013, for a discussion). All of the research assistants were undergraduates. Although the research assistants were familiar with the task prior to the experiment, there was no reason to expect that this familiarity would influence their partners' memory.[2] In our analysis of the development of common ground, the verbal data produced by the research assistants was included in the analyses, but since research assistants did not complete the memory tests, their data did not contribute to the memory measures.

Finally, test data from six participants were dropped: four due to experimenter or computer error, one participant did not follow directions, and one participant was color-blind. In sum, all of the data analyses were conducted on a total of 55 participants who completed the experiment in one of the 18 groups of participants.

[1] In order to evaluate whether English fluency influenced the results, a post hoc analysis coded whether any of the four participants spoke with an accent. In post hoc analyses we added accent as an additional factor into our main analyses; these analyses revealed no significant effects of accent and no interactions.

[2] Additional post hoc analyses were conducted for the subset of groups for which all four participants were true participants; the pattern of results was the same as in the overall analysis.

Materials Visual stimuli were 128 pictures taken from the Internet. The collection of pictures was composed of eight basic object categories of 16 pictures each. Within each category, the pictures were selected to provide 16 clear depictions of objects from that category. Additionally, the pictures in each category were selected to avoid any one of the pictures from being too distinctive from the others within each category. The categories were red leaves, blue and yellow fish, butterflies, drinking containers (i.e., cups, mugs, and glasses), pink flowers, frogs, rabbits, and birds (see Fig. 1 for examples). For each group of participants, eight pictures were randomly selected from each category, with two pictures assigned to each of the four rounds of game-play. The remaining eight pictures from each category served as new items on the subsequent recognition memory test. On the test, this design yielded 64 old pictures and 64 new pictures (which pictures were old vs. new was randomized across groups of participants). Of the pictures that were old, half were pictures that the participant saw as a director, and half were pictures that the participant saw as a matcher. Of the pictures that were seen as a director, half were described to one participant, and half were described to a different participant. Of the pictures that were seen as a matcher, one participant had described half of them, and a different participant had described the other half. Therefore, from the participant's perspective, each picture was paired with only one unique Role × Partner combination. For each picture, there were only two possible choices for which partner could be associated with a given picture, such that if one person served as a source (or destination) in one round, that same person served as a destination (or source) on a later round.

Design The experiment consisted of four rounds of game-play. In the first two rounds, two participants were randomly assigned to be directors and the other two were assigned to be matchers. Those assigned as directors in the first round stayed in the same room throughout the experiment and those assigned to be matchers switched rooms during the experiment (see Table 1). In the first round, Director A interacted with Matcher B (A to B), and Director C interacted with Matcher D (C to D). In the second round, the directors switched matchers such that Director A now interacted with Matcher D (A to D), and Director C now interacted with Matcher B (C to B). In rounds three and four, the roles were switched such that the directors were now matchers, and the matchers were now directors. Half of the time, these new matchers switched partners from round two to three (e.g., C to B in round two, and B to A in round three), and the other half of the time, these matchers did not switch partners from round two to three (e.g. C to B in round two, and B to C in round three). In round four, all the matchers switched partners once more. Across rounds, each participant served as a director with two different participants and as a matcher with those same two participants.

Procedure and equipment Participants completed the conversation task in groups of four individuals. Each person in the group was given a name tag and was addressed by name throughout the experiment. There were two adjacent testing rooms, each with two computers. Each person in the group was assigned to one of the computers in the testing rooms such that each person used their own computer. The computers were situated such that participants were facing each other and could not see each other's display but could see each other's faces.

In cases where there was only one genuine participant in the group, the participant was assigned to be the director for the first round of game-play and one of the research assistants was assigned to be a matcher. This participant remained stationary throughout the experiment while two research assistants switched places each round (as dictated by the experimental design).

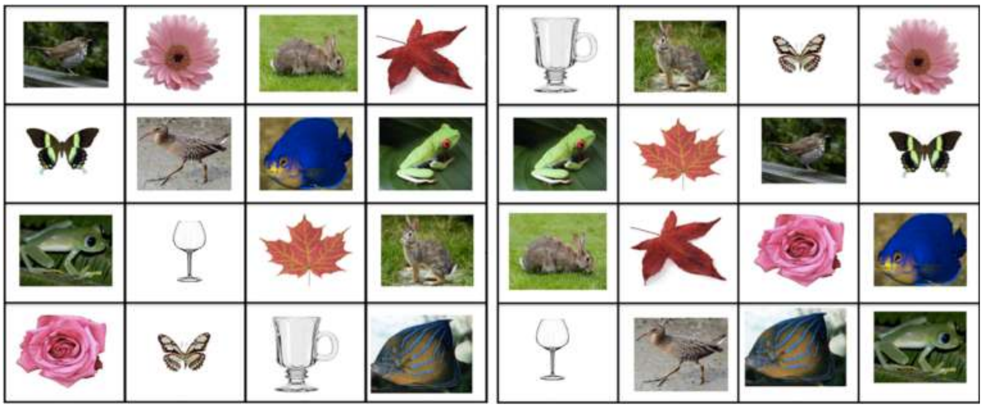

Fig. 1 Example of the director's display (*left*) and the matcher's display (*right*) at the beginning of a given trial. The same set of pictures was used in each trial within each of the four rounds of game play. For each of the three trials within a round, the arrangement of the displays was randomly determined. The director's task was to instruct the matcher on how to rearrange his or her pictures to be in the same order as the director's pictures. (Color figure online)

Table 1 A schematic that details the pairing of directors with matchers in each room for a given group

	Room 1		Room 2	
	Director	Matcher	Director	Matcher
Round 1	A	B	C	D
Round 2	A	D	C	B
Round 3	B[a]	A	D[a]	C
Round 4	D[a]	A	B[a]	C

[a] For half of the groups, the director was in the opposite room.

Each dyad was instructed that, in a given round, each person would see a set of pictures that was identical to their partner's and that the only difference between their displays was the ordering of the pictures (see Fig. 1). The director's task was to describe each picture to the matcher from left to right and from top to bottom so that the matcher could rearrange his or her pictures into the same order. Once all of the pictures had been described and the two participants established that they were finished, they were both instructed to right-click to progress to the next trial where they would repeat the process with the same pictures, but in a different arrangement. Within each round, participants completed three entrainment trials with the same pictures; the purpose of the entrainment trials was for the conversational partners to establish shared labels for each picture. It is the mutual knowledge of the image labels that constitutes the partners' common ground. When it was dictated by the design, the matchers switched rooms with each other after the dyads completed the entire round. On the third round, the matchers of each dyad also switched roles with the directors. An experimenter was present in each room to answer any questions, coordinate the switching of rooms, and to ensure that each partner began and finished each trial at approximately the same time. Although the director and matcher both terminated each trial at roughly the same time, the experimenter was there to communicate when both screens were ready before initiating the appearance of the pictures.

Presentation of the visual images and recordings of both participants' voices were controlled using Psychophysics Toolbox for MATLAB (Brainard, 1997). Audio of the director and matcher's voices was recorded using tabletop microphones over the open sound field. Each trial began with an empty 4 × 4 grid on the screen. Participants used the mouse to left-click their screen, which filled each cell in the grid with a picture. Once the participant right-clicked, the display disappeared, and the computer stopped the audio recording and saved the recording to disk.

After all four rounds of conversation had been completed, participants were given a memory assessment for the information in the experiment. To minimize context effects (Smith & Vela, 1992), each participant was tested in a separate room that was new to that participant.

Participants were then presented with a series of individual pictures. For each picture, participants made two judgments. First, they made an old–new judgment and were asked to press "W" if they remembered seeing the picture during the conversation task and "O" if not. Participants were told that some pictures would be old, and some new, but they were not informed about the relative proportions. Regardless of their response, participants were then asked which partner they had seen the picture with (recall that each participant completed the task with two separate partners, serving in both the director role and the matcher role). The names (e.g., Bob and Sue) of each participants' former partners were presented on the screen, and participants were instructed to press "S" for the name on the left, and "L" for the name on the right. If the participant indicated that the picture was new, they were instructed to guess who they may have seen it with as if the picture was old. This is an established procedure in memory research (e.g., Benjamin, 2006; Benjamin, Diaz, Matzen, & Johnson, 2012; Starns, Hicks, Brown, & Martin, 2008), and was used here to avoid creating any incentive to respond "new" to each item.

We expected that participants would establish brief labels to refer to each picture across the three trials within a round (Wilkes-Gibbs & Clark, 1992). This drop in the number of words used to describe a given picture reflects the development of common ground. Our goal is to relate this process of establishing common ground to memory for the discourse.

If the process of forming common ground promotes memory for what has been discussed, then the reduction in the number of words used by the director over trials should be positively related to item recognition performance. Such a finding would suggest that the act of collaboratively generating labels (e.g., "It looks like a fish that . . . .") and shaping them through repeated reference improves memory for these discourse topics. Similarly, if forming common ground promotes memory for the person with whom you share that common ground, then the reduction in length of the directors' descriptions should be positively related to context memory. That is, if forming and using common ground during conversation depends on the associations between a particular referent and a discourse partner, then the development of common ground should be positively related to accurately identifying the appropriate partner associated with a given picture.

While the formation of common ground is a collaborative, interactive process, each conversational partner leaves the conversation with his or her own personal record of the conversation. Thus, even if the process of forming common ground promotes memory for the conversation, a separate question is whether the speaker and addressee form equivalent memories of what they discussed. If the generation effect extends to unscripted conversation, we should find an egocentric bias in what is remembered, such that speakers will remember their own contributions to the discourse better than their partner's on an item-recognition test. Similarly, if conversational demands unevenly distribute the burden of considering joint knowledge, then one might expect destination and source memory to differ. If the benefits of generation extend to context memory (E. J. Marsh et al., 2001), then we would expect directors to outperform matchers. By contrast, if there are trade-offs between item and context memory (Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; Koriat et al., 1991), matchers may outperform directors for context memory.

Analysis and results

Conversation task During the conversation task, the matchers were highly successful at following the director's instructions, making errors on only 6% of the trials.

The first author and six research assistants transcribed the recordings from the conversation task. Each transcription was checked and corrected for errors by two separate research assistants. The following is an example transcript for the interaction corresponding to two (nonconsecutive) yellow and blue fish (see Fig. 1 for the specific pictures) in the entrainment Trials 1–3 of a given round:

(1) Trial 1

Director: *and then the fish with the blue stripes on it* [word count = 10]
Matcher: *k*
Director: *and then next the other fish the mostly blue one with yellow* [word count = 12]
Matcher: *ok*

(2) Trial 2

Director: *thee blue fish with the yellow face* [word count = 7]
Director: *the fish with the blue stripes* [word count = 6]
Matcher: *ok*

(3) Trial 3

Director: *the fish with the blue stripes* [word count = 6]
Matcher: *ok*
Director: *then the fish that is mostly blue* [word count = 7]

The transcripts were used to count the total number of words used to identify each picture by the director and matcher (see example director word counts above). This total word count included all descriptive words and phrases as well as any lexical disfluencies (e.g., "um," "uh"), location information (e.g., "the top left picture is"), and any kind of feedback from the matcher (e.g., "ok"). The analysis excluded words that did not pertain to locating or identifying any of the pictures (e.g., "are you ready?"); pauses between words were not included in the total word count (e.g., "the red . . . leaf" would count as three words). These measures were intended to be inclusive of all talk needed to accomplish reference, including lexical disfluency (e.g., Clark & Fox Tree, 2002). Utterance length, defined as the total number of words used to identify each picture, was computed separately for directors and matchers for each trial during entrainment.

We used these utterance length data in two ways. First, we analyzed the change in utterance length across the entrainment trials for utterances produced by both directors and matchers. Second, in our primary analyses, we relate memory for the conversation to the development of common ground, using only the length of the director's utterances. This was done because the director must initiate the description of each picture, whereas the matcher may only ask for clarification and may not say anything at all. We assume that matchers will ask for clarification when necessary. Under these assumptions, the directors' utterances would consistently reflect the initial effort that went into describing each image, as well as the efficiency that emerged through the conversational process as common ground was formed (see Clark & Wilkes-Gibbs, 1986, for a discussion of the relationship between utterance length and effort). The choice to restrict our primary analyses to only the director's utterances was corroborated by the fact that utterance length was shorter and varied much less for matchers (M = 1.61, SD = 2.65) than for directors (M = 9.42, SD = 6.14), and that matchers made very few errors. The picture descriptions typically start out quite long, but as partners develop a shared conceptualization of each picture, they typically arrive at shorter labels to describe each picture. We measure the decrement in utterance length from Trial 1 to Trial 3 to assess the development of these shared labels, while taking into account that some pictures may be overall harder to describe than others. Thus, our measure of the process by which conversational partners developed common ground was the difference in the number of words that directors used to describe each picture on Trial 1 and Trial 3 (T1 − T3). In Example 1–3 above, the T1 − T3 measure for the "striped" fish would be 4. Note that while this measure is intended to capture the efficiency gains as common ground accumulates, it necessarily also reflects the initial effort that went into describing each picture (T1). See the Appendix Tables 5 and 6 for analyses that alternatively examine the contribution of effort (captured by the number of words used to describe each picture on T1) to memory.

The data were analyzed with planned, mixed-effects models and a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013) that captured variability across items, subjects, and the nesting of subjects within groups. As a result, variance associated with, for example, item-specific difficulty, would be placed in the random part of the model rather than the fixed part of the model. When modeling common ground development, the linear linking function was used. When modeling item or context memory, the logit linking function was used. For models that did not converge, a backward-fitting approach was used in which the random effect that captured the least variance was removed for each consecutive model until it converged. The models of item and context memory originally would not converge after a substantial number of random effects had been removed. In response, the BOBYQA optimizer was used for these models. When modeling common ground development, significance tests for fixed effects were obtained using a likelihood ratio test in which the full model was tested against a model with the same random effects structure, but without the fixed effect in question. When modeling item or context memory, significance tests for the fixed effects were provided by the model fits that were estimated by the Laplace approximation of the maximum likelihood.

Entrainment trials Analysis of utterance length during entrainment trials was conducted to first establish that interlocutors establish brief, entrained terms for the pictures. Figure 2 displays the average utterance length that was used to describe each picture, as a function of trial and role, collapsed across the four rounds of game-play. The number of words that were used to describe a picture in each trial was centered and modeled using a mixed-effects model with role (director vs. matcher) and trial order as centered fixed effects. As shown in Table 2, there was a significant effect of role such that directors spoke more than matchers. There was a significant effect of trial, indicating that the number of words per picture decreased as a function of trial, consistent with the development of shared labels for the pictures. A Role × Trial interaction was due to the fact that the number of words per picture decreased more for the directors than for the matchers (Fig. 2).

Memory and common-ground development

Item recognition We now turn to the relationship between the development of common ground and item memory. The degree to which directors shortened their labels for a given picture from Trial 1 to Trial 3 (in terms of the number of words) was used as an index of the development of common ground for the picture label (i.e., T1 − T3). On average, utterances from directors decreased by 4.77 words from Trial 1 to Trial 3 ($SD = 7.31$, min = −41, max = 56). This measure was centered and included as a fixed effect, along with participant role to predict memory for the same items (see Fig. 3) . The results of this analysis are shown in Table 3. The analysis was restricted to old items because all of the predictors were only defined for old items. This fact prevents us from looking at whether false alarms differed across role. However, the average false alarm rate was relatively low ($M = 0.107$, $SD = 0.074$, min = 0, max = 0.375). Because the parameter estimates are in log-odds space, effects in odds are computed by taking the exponential of the estimate. There was a significant effect of role, indicating that the odds of recognizing a picture was 2.64 times greater when the participant described the picture in the role of director than when they listened to that description as a matcher. There was a significant effect of common ground development. For each reduction of one word that the director used to describe a picture, the odds of recognizing that picture increased by 1.03 times. The Role × (T1 − T3) interaction was not significant, indicating that this pattern did not differ significantly across directors and matchers.

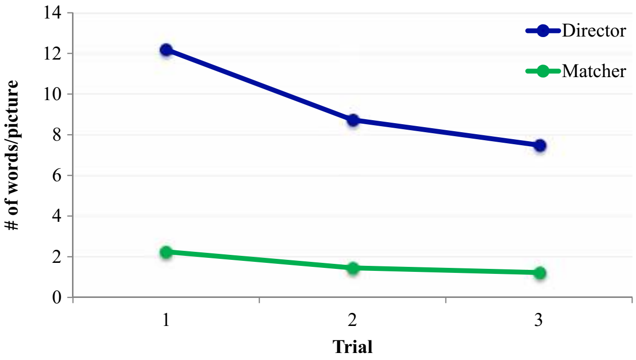

Fig. 2 Utterance length as a function of trial for directors and matchers. (Color figure online)

Context identification To understand the relationship between common-ground development and context memory, a logistic mixed-effects model was used to model the correct context identification with the same fixed effects as the previous analysis (see Fig. 4). As shown in Table 4, the effect of role was not significant, indicating that directors correctly identified the conversational partner that was associated with each picture just as well as matchers did. A significant effect of common-ground development indicated that for each reduction of one word that the director used to describe a picture, the odds of context identification increased by 1.02 times. The interaction was not significant.

Discussion

Participants in an unscripted, task-based conversation developed common ground for a series of picture labels. Directors initially produced long, effortful descriptions of each image that were collaboratively shortened with repeated reference. This process of developing common ground proved beneficial for both item memory (memory for the discussed pictures) and context memory (memory for the partner with whom you discussed those pictures). The significant main effect of common ground formation (T1 − T3) can be thought of as a person-by-item covariate (see Goodwin, Gilbert, & Cho, 2013); the random effects structure of our statistical models allows us to conclude that the observed effects on both item and context memory are above and beyond any contribution

Table 2 Mixed-effects model predicting utterance length during Entrainment

	Estimate	*SE*	*t* value	*p* value			Variance	*SD*
Fixed					*Random*			
Intercept	0.007	0.221	0.030		*Subject*	Intercept	2.420	1.556
Role	7.820	0.409	**19.110**	**<.001**		Role	6.594	2.568
Trial	−1.404	0.102	**−13.790**	**<.001**		Trial	0.450	0.671
Role × Trial	−1.821	0.159	**−11.470**	**<.001**		Role × Trial	0.904	0.951
					Item	Intercept	0.410	0.640
						Role	0.836	0.915
					Group	Role	0.601	0.775

Number of observations = 10,560; number of items = 128; number of subjects = 55; number of groups = 18

of individual items or participants and instead tells us about the way partners worked together to talk about the pictures, and how this relates to memory. In addition, directors outperformed matchers in their memory for these pictures.

Forming common ground

The joint creation of brief labels to refer to each picture served the primary purpose of allowing the conversational partners to quickly and efficiently communicate about the pictures. A large body of research on conversational processes shows that this process of forming common ground, which we index here through a measure of referential shortening (T1 − T3), supports efficient communication (Clark & Wilkes-Gibbs, 1986; Hupet, Chantraine, & Nef, 1993; Wilkes-Gibbs & Clark, 1992). Conversely, in situations where conversational partners cannot form common ground due to an inability to interact, these efficiency gains are much smaller, evidenced by attenuated referential shortening across rounds (Krauss & Weinheimer, 1966). Likewise, when partners lack common ground, for example, when speaking to a new partner who is unfamiliar with the pictures and their labels, speakers tend to revert to long descriptions in order to accommodate the new partner's naïveté (Horton & Gerrig, 2002; Isaacs & Clark, 1987; Schober & Clark, 1989; Yoon & Brown-Schmidt, 2014). Our findings show that the communicative gains that come from establishing common ground are complemented by memorial benefits. The more the director shortened their referential label, the better both conversational partners remembered that picture. Thus, not only does referential efficiency support communication in the moment, it promotes future memory for that conversation as well.

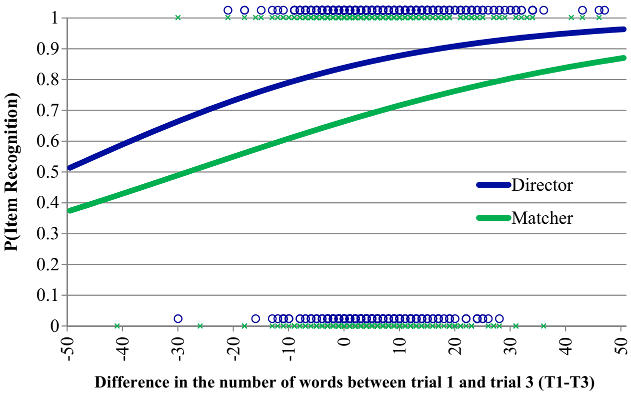

Fig. 3 The probability of recognizing an item as a function of role and the difference in the number of words used by the director from Trial 1 to Trial 3 (T1 − T3). *Solid lines* indicate model fits; response data are shown at the top and bottom of the plot. Blue circles = director responses; green *X*s = matcher responses. (Color figure online)

An important note is that our measure of the development of common ground (T1 − T3) is necessarily influenced by the initial effort that went into describing each picture on the first trial of the task (T1). There was natural variation in how many words were needed to accomplish reference on the first trial, and the length of this initial communicative effort caps the maximum change that could be expected through referential shortening (the correlation between T1− T3 and T1 was 0.834). Thus, while the focus of our a priori analyses are the gains in efficiency seen with the development of common ground across trials, the initial effort that went into establishing common ground (T1) also promoted memory (for a summary, see the Appendix Tables 5 and 6). Indeed, in any natural communicative exchange, there is likely to be a relationship between the length of an initial description, and the amount of shortening with repeated reference. Due to the high degree of collinearity between these measures in our dataset (and quite likely in any natural conversation), it is impossible to parcel out the unique contributions of these two measures with a high degree of certainty. However, it should be noted that any explanation involving differences in items (e.g., picture complexity or distinctiveness) are unlikely given that the random effects in the models explicitly account for such variability. Even with this caveat, the main conclusion from our work, that the process of developing common ground—which includes an initial process of establishing reference, followed by a collaborative referential shortening over time—promotes memory for speakers and listeners alike. It is likely that multiple influences play important roles in enhancing or

Table 3 Mixed-effects model predicting item recognition

	Estimate	*SE*	*z* value	*p* value			Variance	*SD*
Fixed					*Random*			
Intercept	1.180	0.104	**11.353**	**<.001**	*Subject*	Intercept	0.237	0.487
Role	0.969	0.149	**6.509**	**<.001**		Role	0.699	0.836
T1 − T3	0.028	0.009	**3.308**	**<.001**		Role × T1 − T3	0.003	0.056
Role × T1 − T3	0.008	0.017	0.460	.646	*Item*	Intercept	0.509	0.713
						Role	0.036	0.188
						T1 − T3	0.001	0.036
						Role × T1 − T3	0.000	0.017
					Group	Intercept	0.001	0.038
						Role	0.003	0.052
						Role × T1 − T3	0.000	0.015

Number of observations = 3,520; number of items = 128; number of subjects = 55; number of groups = 18

weakening memory for conversation beyond participant role and the formation of common ground, including, for example, interaction style (Pasupathi, Stallworth, & Murdoch, 1998), and retrieval processes (Karpicke & Roediger, 2007).

Conversational role

While the process of forming common ground benefited conversational memory for both speakers and listeners, this finding did not imply that speakers and listeners had equivalent memory for the discourse history. Instead, we found that speakers tended to have better memory for what had been discussed than listeners.

A consistent finding in the memory literature is that generating items out loud tends to boost item memory performance compared to reading the items silently (e.g., MacLeod et al., 2010; Slamecka & Graf, 1978). Although some studies in the memory literature have observed generation to have a negative impact on context memory (e.g., Brown et al., 1995; Fischer et al., 2015; Gopie & MacLeod, 2009; Jurica & Shimamura, 1999; Koriat et al., 1991, Experiment 2), the results for context memory have been mixed (e.g., E. J. Marsh et al., 2001). Here, we replicated the well-known generation effect for item memory (Jacoby, 1978; Koriat et al., 1991; Slamecka & Graf, 1978; Riefer, Chien, & Reimer, 2007) in a naturalistic conversational setting. On the other hand, we found no difference between speakers and listeners in their ability to remember which person they had discussed individual pictures with. Some research suggests that the effect of generation on context memory may depend on specifics of the task (R. L. Marsh & Hicks, 2002) and/or how context is defined (Mulligan, Lozito, & Rosner, 2006; Riefer et al., 2007). A generation penalty on context memory may be more likely in tasks that induce a higher degree of self-focus (Gopie

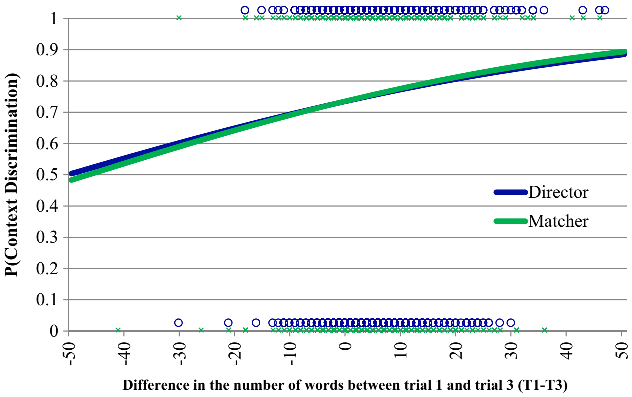

Fig. 4 The probability of identifying the context as a function of role and the difference in the number of words used by the director from Trial 1 to Trial 3 (T1 − T3). *Solid lines* indicate model fits; response data are shown at the top and bottom of the plot. Blue circles = director responses; green *X*s = matcher responses. (Color figure online)

Table 4 Mixed-effects model predicting context identification

	Estimate	*SE*	*z* value	*p* value			Variance	*SD*
Fixed					*Random*			
Intercept	1.031	0.090	**11.444**	**<.001**	*Subject*	Intercept	0.241	0.491
Role	−0.005	0.111	−0.042	.966		Role	0.063	0.251
T1 − T3	0.021	0.008	**2.702**	**.007**		T1 − T3	0.000	0.011
Role × T1 − T3	−0.002	0.015	−0.113	.910		Role × T1 − T3	0.000	0.007
					Item	Intercept	0.159	0.399
						Role	0.002	0.039
						T1 − T3	0.000	0.002
						Role × T1 − T3	0.000	0.018
					Group	Role	0.022	0.148
						T1 − T3	0.000	0.007
						Role × T1 − T3	0.000	0.013

Number of observations = 2,531; number of items = 128; number of subjects = 55; number of groups = 18

& MacLeod, 2009), such as disclosing personal information (Fischer et al., 2015). The lack of a generation penalty in the present research is noteworthy, given the clear generation benefit for item memory and the fact that this is the only experiment that has examined the issue with a conversational task that approximates natural dialogue and a memory task that controls for important response characteristics.

The fact that we observed a speaker benefit for item memory suggests that over the course of a conversation, interlocutors are likely to develop distinct memories of the discourse, such that self-produced utterances are remembered better. The locus of the generation benefit for item memory may be that producing an object description requires more effort than comprehending a description, and that this added effort strengthens the speaker's memory for the item.

Our results emphasize the relevance of findings from traditional memory paradigms—in this case, the generation effect (e.g., Slamecka & Graf, 1978)—to language use in conversational settings. Indeed, our findings align well with previous reports that speakers are more likely to reuse utterances they themselves had produced previously (compared to their partner's utterances; Knutsen & Le Bigot, 2014). More generally, our findings emphasize the importance of examining contributions of memory for the content of a discourse (Ross & Sicoly, 1979) to language use (e.g., Horton & Gerrig, 2005a, 2002). The presence of a generation effect in the current study also highlights the importance of controlling for response variables, like the list-strength effect or output order, when comparing memory between interlocutors, and suggests that the absence (e.g., Knutsen & Le Bigot, 2014) or reversal (e.g., Stafford & Daly, 1984; Stafford et al., 1987) of a generation effect may reflect differences in how people respond or withhold information rather than differences in memory. For example, in a free-recall test of memory for conversation, participants may withhold information about one's own contribution simply because they put more emphasis on reporting their partner's contribution.

It is well established that interlocutors develop coordinated representations of the discourse (Pickering & Garrod, 2004, 2013), and that these joint representations of common ground support understanding (Richardson & Dale, 2005; Richardson et al., 2007; Schober & Clark, 1989). The present findings show that even during an interactive conversation where common ground is being actively established, basic memory effects, such as the mnemonic benefits of generation, will limit the degree to which coordination is possible. While establishing common ground confers memorial benefits to both speakers and listeners, speakers are likely to remember their contributions to the conversation better, preventing perfect coordination. Our findings suggest that contributions to future interactions will be driven by what each person had contributed in the past (see Knutsen & Le Bigot, 2012, for a similar discussion), and that representations of common ground may become disproportionate over longer periods of time.

The fact that item detection was higher for speakers than listeners and that context identification did not differ between speakers and listeners points to an interesting implication of these findings. One may be equally likely to remember the destination of an utterance as to remember the source. However, because one is more likely to remember an utterance that they produced, they may become more aware of subsequent memory failures that involve identifying the appropriate context. In contrast, if one is less likely to remember an utterance that they heard, then they may be unaware of the fact that they also do not remember the source of that information. As a result, people may overestimate the frequency with which they forget the destination of a previous utterance because they have better memory for the utterance itself.

In conclusion, the process of developing common ground for referential labels promoted both item and context memory.

Thus, not only does forming common ground support efficient communication in the present, the process of forming common ground promotes a more firmly established memorial record of the discourse history. While both speakers and listeners benefited through the process of forming common ground, the observation that speakers outperformed listeners suggests that perfect alignment between conversational partners will be frequently out of reach, even in interactive conversational settings. This finding emphasizes the fact that even when interlocutors are cooperatively forming common ground, differences in memory for the discourse are likely. Thus, while forming common ground is an inherently cooperative enterprise, when each individual steps away from that conversation, they each hold onto a separate, imperfect, record of what they shared. We show that the process of forming common ground improves the memorial record of our communicative exchanges, but that these memories are asymmetric.

Acknowledgements This material is based on work supported by National Science Foundation Grants NSF BCS 1257029 and BCS 1556700 to Sarah Brown-Schmidt. We would like to thank Michael Agapito, Phoebe Bauer, Jessica Chu, Jensen Davis, Nathan Couch, and Taylor Salata for assistance with data collection and coding.

Appendix

Table 5 Mixed-effects model using role and utterance length from each trial to predict item recognition

	Estimate	*SE*	*z* value	*p* value			Variance	*SD*
Fixed					*Random*			
Intercept	1.242	0.111	**11.241**	**<.001**	*Subject*	Intercept	0.241	0.491
Role	0.998	0.143	**6.983**	**<.001**		Role	0.583	0.764
T1	0.050	0.011	**4.505**	**<.001**		T1	0.000	0.013
T2	0.006	0.014	0.412	.680		T2	0.001	0.025
T3	0.055	0.020	**2.770**	**.006**		T3	0.000	0.016
					Item	Intercept	0.595	0.771
						Role	0.032	0.178
						T1	0.002	0.044
						T3	0.002	0.045
					Group	Role	0.009	0.093
						T1	0.000	0.016
						T2	0.001	0.030
						T3	0.002	0.042

Table 6 Mixed-effects model using role and utterance length from each trial to predict context identification

	Estimate	*SE*	*z* value	*p* value			Variance	*SD*
Fixed					*Random*			
Intercept	1.081	0.094	**11.458**	**<.001**	*Subject*	Intercept	0.207	0.454
Role	0.008	0.119	0.066	.948		Role	0.019	0.137
T1	0.026	0.008	**3.206**	**.001**		T2	0.001	0.031
T2	0.005	0.014	0.362	.718		T3	0.001	0.031
T3	0.014	0.021	0.680	.496	*Item*	Intercept	0.161	0.402
						Role	0.003	0.054
						T2	0.000	0.022
						T3	0.005	0.072
					Group	Intercept	0.014	0.119
						Role	0.063	0.251
						T2	0.000	0.022
						T3	0.002	0.048

References

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74,* 81–99.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46,* 391–418.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review, 12,* 874–879.

Benjamin, A. S., Diaz, M., Matzen, L. E., & Johnson, B. (2012). Tests of the DRYAD theory of the age-related deficit in memory for context: Not about context, and not about aging. *Psychology and Aging, 27*(2), 418–428.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Social Psychology: Learning, Memory, and Cognition, 22,* 1482–1493.

Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational source monitoring. *Psychology & Aging, 10,* 111–122.

Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61,* 171–190.

Brown-Schmidt, S. (2012). Beyond common and privileged: Gradient representations of common ground in real-time language use. *Language Cognitive Processes, 27,* 62–89.

Brown-Schmidt, S., & Fraundorf, S. H. (2015). Interpretation of informational questions modulated by joint knowledge and intonational contours. *Journal of Memory and Language, 84,* 49–74.

Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine, & S. D. Behrend (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84,* 73–111.

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

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science, 13,* 259–294.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1–39.

Egan, J. P. (1958). *Recognition memory and the operating characteristic (Tech. Note AFCRC-TN-58–51)*. Bloomington: Indiana University, Hearing and Communication Laboratory.

Fischer, N. M., Schult, J. C., & Steffens, M. C. (2015). Source and destination memory in face-to-face interaction: A multinomial modeling approach. *Journal of Experimental Psychology: Applied, 21*(2), 195.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2015). Reduction in prosodic prominence predicts speakers' recall: Implication for theories of prosody. *Language, Cognition and Neuroscience, 30*(5), 606–619.

Goodwin, A. P., Gilbert, J. K., & Cho, S. J. (2013). Morphological contributions to adolescent word reading: An item response approach. *Reading Research Quarterly, 48*(1), 39–60.

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 expression reflects shared experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 843–853.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13,* 512–521.

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 knowing *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.

Hupet, M., Chantraine, Y., & Nef, F. (1993). References in conversation between young and old normal adults. *Psychology and Aging, 8,* 339–346.

Isaacs, E., (1990). *Mutual memory for conversation* (Doctoral dissertation, Stanford University, Standord, CA).

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

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17,* 649–667.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information: Evidence for a negative generation effect in source memory. *Memory & Cognition, 27,* 648–656.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 57,* 151–162.

Keenan, J. M., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study in natural conversation. *Journal of Verbal Learning and Verbal Behavior, 16,* 549–560.

Knutsen, D., & Le Bigot, L. (2012). Managing dialogue: How information availability affects collaborative reference production. *Journal of Memory and Language, 67,* 326–341.

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review, 21,* 1590–1599.

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of memory for input and output events. *Psychological Research, 53,* 260–270.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517.

Krauss, R. M., Garlock, C. M., Bricker, P. D., & McMahon, L. E. (1977). The role of audible and visible back channel responses in interpersonal communication. *Journal of Personality and Social Psychology, 35,* 523–529.

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.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review, 20,* 54–72.

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.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah: Erlbaum.

MacWhinney, B., Keenan, J. M., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory & Cognition, 10,* 308–317.

Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & Cognition, 29,* 798–805.

Marsh, R. L., & Hicks, J. L. (2002). Comparisons of target output monitoring to source input monitoring. *Applied Cognitive Psychology, 16,* 845–862.

Miller, J. B., deWinstanley, P., & Carey, P. (1996). Memory for conversation. *Memory, 4,* 615–631.

Mulligan, N. W., Lozito, J. P., & Rosner, Z. A. (2006). Generation and context memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 836–846.

Pasupathi, M., Stallworth, L. M., & Murdoch, K. (1998). How what we tell becomes what we know: Listener effects on speaker's long-term memory for events. *Discourse Processes, 26,* 1–25.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27,* 169–225.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163–178.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29*(6), 1045–1060.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological Science, 18*(5), 407–413.

Riefer, D. M., Chien, Y., & Reimer, J. F. (2007). Positive and negative generation effects in source monitoring. *The Quarterly Journal of Experimental Psychology, 60,* 1389–1405.

Roediger, H. L., & Schmidt, S. R. (1980). Output interference in the recall of categorized and paired associate lists. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 91–105.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37,* 322–336.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21,* 211–232.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4,* 592–604.

Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology, 6,* 125–139.

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory: The effects of time, recall mode, and memory expectancies on remembrances of natural conversations. *Human Communication Research, 14,* 203–229.

Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communication Research, 10,* 379–402.

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition, 36,* 1–8.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language, 31,* 183–194.

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