COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12964

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12964

Referential Form and Memory for the Discourse History

Si On Yoon,[a] Aaron S. Benjamin,[b] Sarah Brown-Schmidt[c]

[a]*Department of Communication Sciences and Disorders, University of Iowa*
[b]*Department of Psychology, University of Illinois at Urbana-Champaign*
[c]*Department of Psychology and Human Development, Vanderbilt University*

Received 24 July 2019; received in revised form 14 February 2021; accepted 24 February 2021

Abstract

The way we refer to things in the world is shaped by the immediate physical context as well as the discourse history. But what part of the discourse history is relevant to language use in the present? In four experiments, we combine the study of task-based conversation with measures of recognition memory to examine the role of physical contextual cues that shape what speakers perceive to be a part of the relevant discourse history. Our studies leverage the *differentiation effect*, a phenomenon in which speakers are more likely to use a modified expression to refer to an object (e.g., dotted sock) if they had previously described a similar object (e.g., striped sock) than when they had not described a similar object. Two physical cues—the background that framed the to-be-described pictures and the position of the pictures in the display—were manipulated to alter perceptions about the relevant discourse context. We measured the rate with which speakers modify referring expressions to differentiate current from past referents. Recognition memory measures following the conversation probed what was and was not remembered about past discourse referents and contexts. Analysis of modification rates indicated that these contextual grouping cues shaped perceptions about the relevant discourse context. The contextual cues did not affect memory for the referents, but the memory for past referents was better for speakers than for listeners. Our findings show that perceptions about the relevant discourse history are a key determinant of how language is used in the moment but also that conversational partners form asymmetric representations of the discourse history.

Keywords: Context; Discourse history; Language production; Lexical differentiation; Memory; Recognition; Reference

1. Introduction

In conversation, speakers and listeners exchange information through words, actions, and gestures (Brennan, Galati, & Kuhlen, 2010; Clark, 2016; Hilliard, O'Neal, Plumert, &

Correspondence should be sent to Si On Yoon, Department of Communication Sciences and Disorders, University of Iowa, Iowa City, IA 52242, USA. Email: sion-yoon@uiowa.edu

Cook, 2015). As a conversation progresses, the conversational partners' memory of what has been said is continuously updated. At any given moment in time, each partner's memory for the conversational history constitutes their individual record of the discourse history up to that point. Successful communication critically depends on tailoring what is said so that it is comprehensible given the discourse history (Gordon, Grosz, & Gilliom, 1993; Marslen-Wilson, Levy, & Tyler, 1982; Poesio, Di Eugenio, & Keohane, 2002). In this sense, the discourse history, as well as the immediate context, is a contextual force that shapes what is said (Brennan & Clark, 1996). Theoretical and computational approaches to discourse posit that discourses are functionally structured based on features such as sentence structure, discourse topic, actions, and goals (e.g., Grosz & Sidner, 1986; Ito & Speer, 2008; Tetreault & Allen, 2003).

Despite general agreement that the discourse history shapes language use, very little is known about how the discourse history is encoded in memory for speakers and listeners, and the factors that determine assessments of which aspects of the discourse history are relevant in the moment—the *relevant discourse context*. The aim of the present research is to understand how perceptions of and memory for the discourse context influence language use. The historical discourse context is evanescent and therefore only accessible through memory. But, as we shall see, it impacts language use in much the same way that the immediate context does (Brennan & Clark, 1996).

It is well established that the immediate (or local) discourse context has a strong influence on referring expressions (Olson, 1970). For example, if there are two growlers of porter on the table, a rye porter and a bourbon barrel porter, the expression "the porter" would not suffice to achieve reference. Instead, the speaker would have to use a more specific expression to make her intentions clear to the addressee; for example, "*Gimme the rye porter*" (Olson, 1970; Osgood, 1971; Pechmann, 1989). Classic work examining contextual influences on language shows that the number of candidate referents in the discourse context shapes syntactic parsing (Altmann & Steedman, 1998) and that such context manipulations can be instantiated by the number of visually present characters in scene (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Trueswell, Sekerina, Hill, & Logrip, 1999). These candidate referents affect processing regardless of whether those objects are introduced or "named" in the discourse (Ryskin et al., 2019). That said, not all of the items in the immediate physical context shape referential form—it is the subset of items that are in the relevant context or "referential domain" that guide language use and understanding (Chambers, Tanenhaus, Eberhard, Carlson, & Filip, 2002; Landragin, 2006). In a task-oriented dialog, factors that determine whether something is or is not in the relevant context include whether the item is visible to both speaker and listener as well as whether it is task relevant (Brown-Schmidt & Tanenhaus, 2008; Chambers, Tanenhaus, & Magnuson, 2004; Heller, Parisien, & Stevenson, 2016; Nadig & Sedivy, 2002). If a context item is task relevant, speakers are more likely to design a referring expression that distinguishes the intended referent from that item in the context (Beun & Cremers, 1998; Brown-Schmidt & Tanenhaus, 2008).

By contrast, the historical discourse context refers to what has been discussed in the past by the conversational partners, including past referents, how they were described, and with

whom (Heller & Chambers, 2014; Marslen-Wilson et al., 1982; Wilkes-Gibbs & Clark, 1992). For example, Brennan and Clark (1996) show that in a referential communication task, that if a speaker had previously needed to use a more specific expression to distinguish an intended referent from others in the context (e.g., *the dress shoe*, to distinguish a dress shoe from a loafer and a high heel), that even when the context changes and the dress shoe is the only visible shoe, speakers often persist in using the more specific expression. They argue that considerations of informativeness of referring expressions must be interpreted with respect to a historical model of the discourse history. The concept of a shared discourse history figures prominently into the notion of *common ground* or mutual knowledge (Clark, 1992; Stalnaker, 2002), and specifically into the idea that the common ground grows as conversational partners share information in conversation (Clark & Marshall, 1978). Changes to the discourse context—most notably changes that involve talking to a new partner—have long been known to change referring patterns, with speakers not assuming that the new partner knows about the previously established referential labels (Wilkes-Gibbs & Clark, 1992). Just as physically copresent information can be taken to be in common ground, information can also be taken to be part of common ground when it was *previously* physically copresent, even if it was not explicitly talked about (Wolter, Gorman, & Tanenhaus, 2011; also see Altmann, 2004; Chambers & San Juan, 2008). Because representations of previously experienced information necessarily depend on memory—and the relevant memories fade with time—the historical discourse context is hypothesized to have a weaker influence on language use (Clark & Marshall, 1978).

The idea that cognitive processes shape the ability to determine which aspects of the discourse history (often, which labels) are shared with which partner was fleshed out by Horton and Gerrig (2002, 2005). They showed that when semantic categories served as memory cues for who-knows-what, speakers were more successful at designing language based on the discourse history shared with the current addressee. That is, if partner A is associated with "frogs" and partner B is associated with "pigs," speakers use this category cue (frog vs. pig) as a memory cue to remember which addressee knows what. Research in the memory literature shows that the context in which information is processed plays an important role in how that information is encoded and subsequently retrieved from memory (e.g., Godden & Baddeley, 1975; Smith & Vela, 2001). In the present study, we ask if the context against which potential discourse referents are experienced shapes processing of and memory for those potential discourse referents—in turn shaping language use. We use a contextual cue that, based on prior work in the memory literature, is known to act as an effective reminder of what one has experienced in the recent past. Specifically, that work shows that a background scene that frames a stimulus can act as a cue to remember *other* stimuli that also appeared with that same background scene (Tullis, Braverman, Ross, & Benjamin, 2014). We ask if similar contextual cues can affect perceptions of the discourse context, linking referents in the immediate discourse context with referents in prior contexts when they are perceptually linked through contextual grouping cues such as a background scene.

In prior research (Yoon, Benjamin, & Brown-Schmidt, 2016), we examined a referential phenomenon—*lexical differentiation*—that depends on the discourse history (Van Der Wege, 2009; Yoon, 2018). Consider a case in which a speaker refers to a sock in a situation where

the immediate context contains a (striped) sock and three unrelated items. Then, later in the conversation, the speaker refers to a new (dotted) sock in a context that includes that new sock along with three other unrelated items. In two experiments, we found that speakers use a modifier to describe the second referent; for example, "*the dotted sock*" 23% of the time, significantly more often than if they had not previously described a sock in the conversation (17%; Yoon et al., 2016). A third experiment in Yoon et al. (2016) replicated this finding in cases where the first sock was presented along with a different (e.g., argyle) sock, prompting the speaker to describe this first sock with an adjective (e.g., "striped sock"). When later presented with the new (dotted) sock, speakers again were more likely to modify (24.7%) compared to a case where they had not previously described a sock (16.3%). Although this lexical differentiation effect is small (∼8% increase in the modification rate, compared to modification rates well over 90% when both items are in the immediate context), it has been reported consistently in a series of studies (Van Der Wege, 2009; Yoon & Brown-Schmidt, 2013; Yoon et al., 2016).

The lexical differentiation effect reflects the influence of the discourse history on language use in the moment. One possible explanation for why the effect is small is that speakers have a poor memory for the relevant prior discourse. To explore this hypothesis, Yoon et al. (2016) measured whether the speaker's memory for past referents could be used to predict lexical differentiation. Across three studies, recognition memory for the discourse following the conversation task exhibited no relationship with the apparent influence of the discourse history during the conversation. In other words, whether or not speakers correctly remembered the previous discourse referent (e.g., the striped sock) was not predictive of whether or not speakers used a modifier to describe the second referent (e.g., the "dotted" sock). We also found that the delay between the naming of the prior discourse referent (e.g., the striped sock) and the current discourse referent (e.g., the dotted sock) did not influence modification. Yoon et al. (2016) also reported that speakers had better memory for past referents than listeners, a finding closely related to the generation and production effects (Jacoby, 1978; MacLeod, Gopie, Hourihan, Neary, & Ozbuko, 2010; Slamecka & Graf, 1978). These effects refer to a pair of well-established phenomena in which item memory is superior when those items are generated (e.g., actively generating a word from its first few letters) or produced (e.g., reading a list of words aloud) than when the words are passively received (e.g., listening to a list of words).

In sum, although memory for the discourse history is necessary in order for the differentiation effect to manifest, these findings suggest that the small size of the differentiation effect is not likely due to a failure to remember the discourse history. In other words, memory for past referents is necessary, but not sufficient, for speakers to design language with respect to the discourse history. In the present work, we explore the alternative possibility that speakers did not view the discourse history as *relevant*.

More specifically, one explanation for why the effect of lexical differentiation was small, and not related to measures of later memory, is that speakers may have perceived past referents to be in a meaningfully different discourse context (Grosz & Sidner, 1986), making those past referents inaccessible to or irrelevant for current discourse purposes. If so, speakers may have the capacity to *remember* the past discourse referents when explicitly motivated but

not use that information when designing subsequent referring expressions. It is well known from research on directed forgetting that information can be held in a state in which it can be recognized but does not affect ongoing processing (Elmes, Adams, & Roediger, 1970). The current studies generalize this logic from ongoing learning (Reitman, Marlin, Bjork, & Higman, 1973) to language production.

2. Overview of Experiments 1–4

This paper presents the results of four experiments that test the hypothesis that the discourse history shapes language use in the moment *when the speaker perceives that discourse history to be relevant to the current discourse context*. To test this hypothesis, we use the phenomenon of lexical differentiation—in which speakers design referring expressions to distinguish current from past referents—and manipulate factors hypothesized to alter the perceptions about the relevant discourse context. We hypothesize that differentiation effects will be magnified when speakers perceive past referents to be part of the current discourse context.

In each experiment, participants describe a series of pictures to a conversation partner. In referential communication tasks, the discourse context is largely determined by the visual displays that participants view across the trials. Across the four experiments, we used physical contextual cues in an attempt to shape the speaker's perceptions about what constitutes the relevant discourse history. Measures of memory following the conversation were used to assess the role of memory for past discourse contexts in this process. These measures also afford the opportunity to replicate and extend previous findings that speakers and listeners form distinct memories of the discourse history. In Experiment 1, we simultaneously manipulated two contextual cues that we hypothesized would shape perceptions about what information was a part of the relevant discourse context. The two cues were the color of the background that framed the to-be-described pictures and the position of the pictures in the display. The manipulation is analogous to the way in which physical spaces cue the type and location of items or events within that space. For example, bakeries are likely to contain a variety of products, like cupcakes and bread, and those items are likely to appear at consistent locations across days within the display case. In Experiment 2, we manipulated discourse contexts using salient background pictures and picture position. In Experiment 3, we manipulated these two physical cues—background pictures and picture position— independently. Although the discourse context manipulations in Experiments 1–2 were successful in influencing how people referred to objects, the independent manipulation of backgrounds and picture positions in Experiment 3 each failed to reach significance. Experiment 4 was a direct replication of Experiment 2 in which we aimed to demonstrate the reproducibility of the discourse context effects observed in Experiment 2. Taken together, the experiments clearly demonstrate that the contextual grouping cues shape perceptions about the relevant discourse context and affect how language is used.

3. Experiment 1

3.1. Method

3.1.1. Participants

Seventy-two undergraduates at the University of Illinois at Urbana-Champaign participated in the experiment in return for partial course credit or cash payment ($10). Participants were native speakers of North American English with normal hearing and normal or corrected-to-normal vision. Participants completed the experiment in pairs (36 pairs total).

3.1.2. Materials and procedure

The experiment consisted of two phases: a referential communication task followed by an unexpected memory test. Participants performed an unrelated task (e.g., solving math problems) for 20 min between the referential communication task and the memory test to avoid ceiling performance on the memory test (Yoon et al., 2016). The experiment lasted approximately 1 h.

3.1.3. Referential communication task

The two participants were seated at separate computers and randomly assigned to the roles of speaker and listener. On their respective computer screens, the participants viewed a grid containing three rows and five columns (shown in Fig. 1). On each trial, four pictures were revealed to the speaker and listener, and the speaker's task was to describe the object in a highlighted red box for the listener to click (speakers and listeners were informed that they would always see the same images; the red box was not shown to listeners). After the listener clicked the object on her screen, the speaker clicked to proceed to the next trial. The task was interactive; both participants were allowed to freely talk to each other and encouraged to ask clarification questions if needed.

The 3 × 5 grids were designed such that each of the 15 squares on the grid could contain an image; however, on each trial, only four of these images were revealed to the speaker and listener. Trials were administered in sets of eight, such that participants completed eight trials with a given set of 15 pictures, and then moved onto a new set of 15 images for eight trials, and so on. A schematic of the procedure and sample images are shown in Fig. 2. When one set ended, the next set of eight trials began, with no breaks between sets. Each set of eight trials was designed to include one "setup" trial, one critical "test" trial, and six interspersed filler trials.

We manipulated two factors in a 2 × 2 experimental design: contextual cues (between-subjects) and differentiation (within-subjects). In referential communication tasks such as this one, the discourse context is based on the series of images that the conversational partners view and describe over a series of trials. We manipulated contextual cues to create the perception that a series of trials were part of the same discourse context or not. In the context-cued condition, a distinct background color was used to group the eight trials in each set and distinguish them from the other sets of trials; in addition, the 15 pictures in that set remained

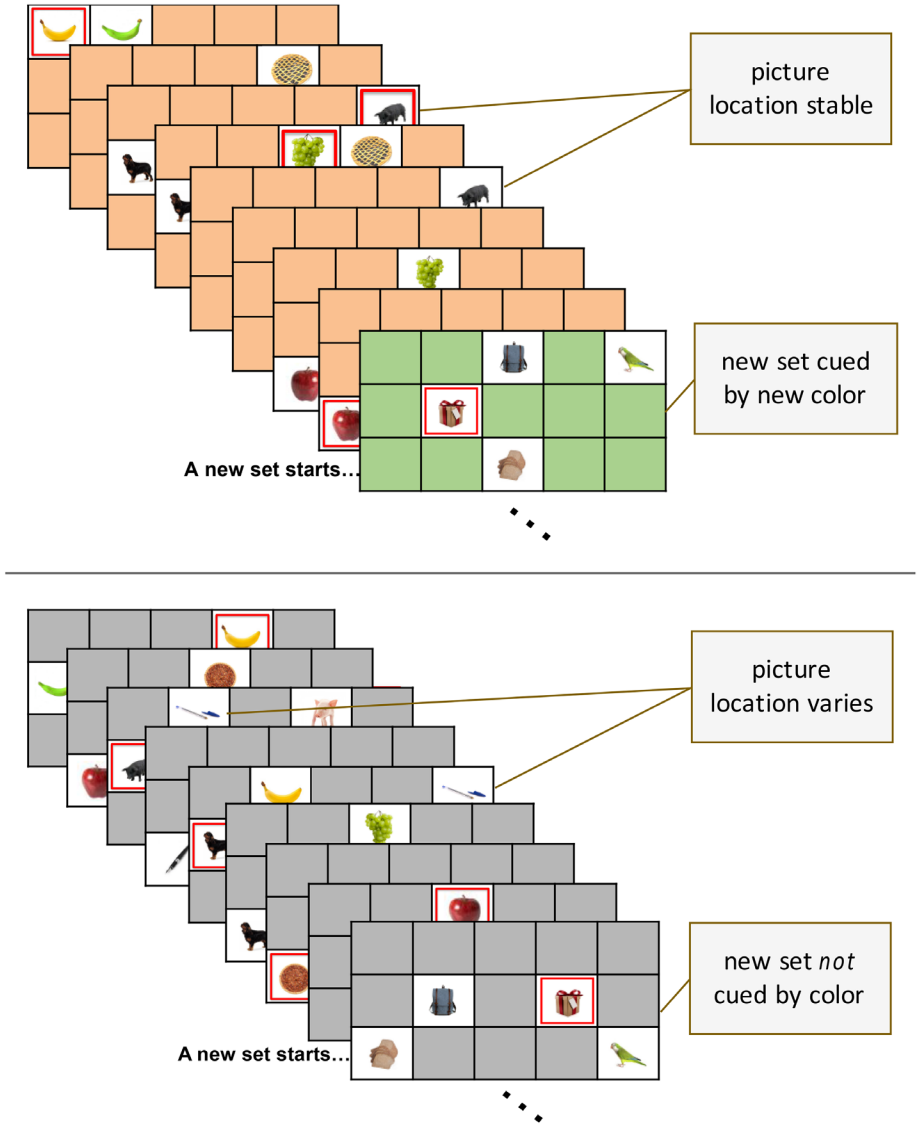

Fig. 1. Experiment 1: Example trial sequence in the context-cued condition (top) and non-cued condition (bottom) during the referential communication task. In the context-cued condition, each set was distinguished by a different background color, and the pictures remained in the same positions across the eight trials within the set. In the non-cued condition, all sets were presented with a gray background, and picture position was randomized.

in fixed grid positions across the eight trials (Fig. 1). The displays in this condition are similar in spirit to a case where participants view a physical display with 15 cubbyholes, where the contents of only four cubbyholes are revealed on any given trial. Across the eight trials within each set, the participants view all 15 images (four at a time). The critical contrast, context, and target items are revealed once, but other filler items could be revealed multiple times. In

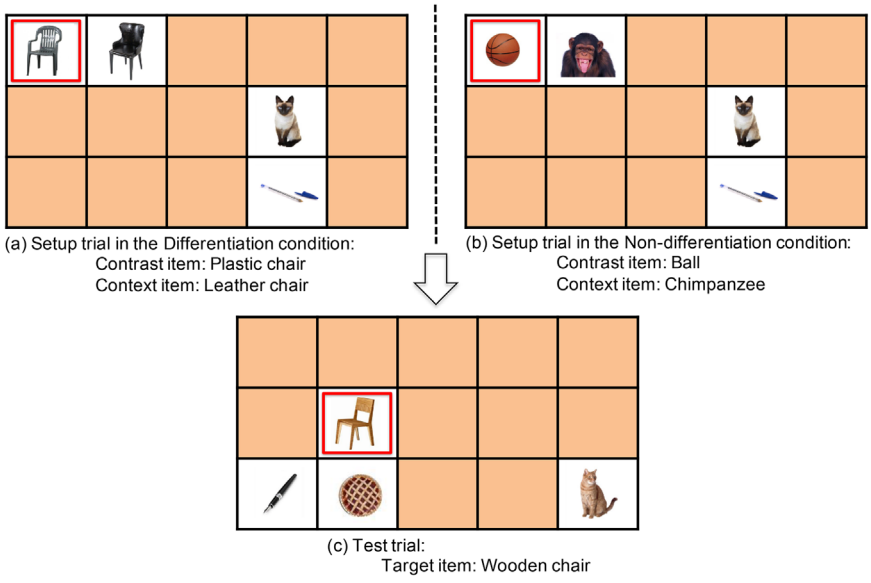

Fig. 2. Experiment 1: Example stimuli in the context-cued condition, illustrating the setup trial in (a) the differentiation condition, (b) the non-differentiation condition, and (c) the test trial. The to-be-named item is indicated to the speaker by the red box. The listener's screen has the same four pictures without the red box. In each set of eight trials, six filler trials were interspersed with the setup and test trials for that set.

the non-cued condition, the background color remained gray for all sets of trials such that one set could not be distinguished from the next; in addition, the position of the 15 pictures within each set was varied randomly.

The differentiation manipulation concerned the nature of the setup trials (Fig. 2). In the differentiation condition, the setup trial had a critical pair of items from a common category (e.g., a plastic and a leather chair), and the speaker described one of them (e.g., "click on the plastic chair"). We use the term *contrast* item for the item that is mentioned on the setup trial (e.g., plastic chair) and *context* item for the unmentioned item in the setup trial that is from the same category as the referent (e.g., leather chair). In the non-differentiation condition, the critical setup pair (e.g., the plastic and leather chair) were never seen by the participants and instead were replaced by a pair of unrelated items (e.g., a ball and a chimpanzee). In this condition, the mentioned *contrast* item was unique, and thus no modifier was necessary (e.g., "click on the ball").

Test trials were the same across all conditions: Speakers described a *target* (e.g., a wooden chair) that was unique in the local context. In the differentiation condition, the target was a third exemplar from the same category (e.g., a third chair); in the non-differentiation condition, the target was the same (e.g., a wooden chair), but the participants had not previously seen or described an object from this category.

There were a total of 28 different item sets, each of which included one setup trial, one test trial, and six filler trials (totaling 28 setup trials, 28 test trials, and 168 filler trials across

the entire study). Within each set, the three trial types were ordered such that the setup trial always preceded the test trial by between zero and six filler trials (average = 3.1). Each set contained three critical items: a contrast and a context item and a target item. The critical items were counterbalanced across conditions through the use of 12 experimental lists. Note that the setup items in the non-differentiation condition (e.g., ball and chimpanzee) were also counterbalanced across lists (sometimes the ball was the referent, and sometimes the chimpanzee was the referent); however, the ball and the chimpanzee never appeared in the differentiation conditions. Each participant was randomly assigned to a single list. The participants switched the roles of speaking and listening midway through the experiment. Thus, each participant completed 14 item sets as a speaker and 14 sets as a listener. The speaker's voice was recorded and saved directly to the computer.

3.1.4. Recognition memory test

Following the communication task, the participants performed an unrelated task (math problems) for 20 min. Participants then completed an unexpected yes/no recognition memory test. Because the memory test took place at a later time than the discourse task, it is likely that forgetting will cause memory for the materials to be poorer than if the two tests had been administered simultaneously. Separating the two tasks was necessary for several reasons. First, recognition memory for pictures immediately after learning would be so near the ceiling (e.g., Shepard, 1967) as to obscure differences between conditions. Second, interleaving memory tests within the discourse would interrupt the discourse and compromise the context manipulation. Third, those tests inform the subject of the centrality of memory to the experiment and might cause them to reactively encode the materials differently in expectation of a test.

Though forgetting is likely to reduce performance, it is unlikely to interfere with our attempts to measure condition effects on memory. At a 20-min delay, performance will still be quite high. More importantly, forgetting rates mostly do not differ as a function of experimental conditions (e.g., Siler & Benjamin, 2020; Slamecka & McElree, 1983) or at least not much (Carpenter, Pashler, Wixted, & Vul, 2008). This means that differences in memory on a later test are a strong indicator of *initial* differences in learning, which is our inferential goal here. In sum, by using a delayed memory test, we increase the chances of detecting meaningful variability in how items were encoded that can then be related to the linguistic phenomena of interest.

On each trial in the memory test, a picture was presented on the screen, and participants were asked to judge whether they had seen the picture during the referential communication task or not. They pressed the "Y" key if they believed the test picture to be one that they had seen during the earlier task and pressed the "N" key if they believed the picture to be new. Half of the pictures were of old items and the other half were of new, previously unseen, items. The old items included 28 contrast objects that were described on a setup trial (e.g., plastic chair or ball), 28 context objects that were present but not described in the setup trial (e.g., leather chair or chimpanzee), 28 targets (e.g., wooden chair), and 28 unrelated filler items (e.g., cookie). The 112 new items were from the same category as each old item

(e.g., in the contrast condition, three new chairs; in the non-contrast condition, a new ball, a new chimpanzee, and a new chair). Old and new items were counterbalanced across experimental lists. The 224 trials were presented in a random order, and participants individually performed the task on their own computer. Participants were not allowed to talk to each other during the memory test, though they were in the same room.

3.2. Predictions

Based on previous findings (Van Der Wege, 2009; Yoon & Brown-Schmidt, 2013; Yoon et al., 2016), we expected speakers to modify their referring expressions; for example, "*wooden chair"* more often when a different exemplar from the same category had been referenced before (differentiation condition) than when no item from that category had been referenced (non-differentiation condition). Using this phenomenon of *lexical differentiation* as a metric, our primary question is whether the context against which potential discourse referents are experienced shapes perceptions of what constitutes the relevant discourse context. If the context manipulation shapes whether speakers view the discourse history as relevant, the differentiation effect should be larger in the context-cued than the non-cued condition. Alternatively, whether or not past referents are considered to be part of the current discourse context may simply be determined by the passage of time or the number of intervening referents (cf. Brown, Neath, & Chater, 2007). If so, the contextual cue manipulation may matter little, and the differentiation effect should be of a similar magnitude in the context-cued and the non-cued conditions. Such an outcome might be expected if past referents remain in the representation of the current discourse context until they are replaced by new incoming information (see Grosz & Sidner, 1986; Kotovsky, Hayes, & Simon, 1985; Newell & Simon, 1972; Walker, 1998). If so, previously mentioned entities may be considered part of the discourse context until a sufficient number of intervening referents has been mentioned or until enough time has passed.

Data from the recognition memory test are used to further probe prior findings that speakers exhibit better memory for past referents than listeners but equivalent memory for unmentioned items in the context (Yoon et al., 2016). Based on those findings, speakers should have better memory for target and contrast items (both of which had been referenced in the conversation) than listeners, but this speaking benefit should not extend to the previously unmentioned (but viewed) context items. Finally, if the contextual cue manipulation does enhance the lexical differentiation effect, the underlying mechanism may be superior *memory* for past referents when they were perceived as contextually linked to other trials. If so, memory for past referents (target and contrast) and context items should be better in the context-cued condition than in the non-cued condition. However, recall that Yoon et al. (2016) found that memory for past referents *did not* predict the differentiation rate. If so, the contextual cue manipulation may affect referential form independently of memory for the discourse history, in which case the contextual cues will increase differentiation without affecting memory. Such an outcome would point to perceived relevance rather than memory as a primary determinant of if and when speakers design language with respect to the discourse history.

Table 1
Proportion of modified expressions (standard deviation of the by-participant means) at test trials during the referential communication task in Experiments 1–4. The effect size (*d*) for the differentiation effect in each condition is shown to the right

	Contextual Cuing		*Non-differentiation*	*Differentiation*	*Effect Size Between the Two Conditions*
Experiment 1	Non-cued		39.9% (26.0)	47.2% (30.3)	0.26
	Context-cued		30.9% (27.9)	47.0% (33.0)	0.52
Experiment 2	Non-cued		36.5% (33.1)	46.8% (29.7)	0.33
	Context-cued		32.5% (26.8)	54.4% (30.2)	0.77
Experiment 3	Gray background	Random	34.1% (30.2)	42.8% (34.1)	0.27
		Item fixed	33.6% (31.7)	44.9% (34.5)	0.34
	Unique background	Item random	30.5% (31.9)	46.8% (35.2)	0.48
		Item fixed	33.5% (37.2)	50.9% (34.0)	0.49
Experiment 4	Non-cued		32.5% (29.0)	39.9% (30.3)	0.25
	Context-cued		30.3% (25.2)	47.4% (31.4)	0.60

3.3. Results

3.3.1. Referential communication task

We transcribed the speakers' referring expressions and coded whether their expressions were modified or not. The modification rates across Experiments 1–4 are summarized in Table 1. In Experiment 1, speakers produced modified noun phrases 47% of the time in the differentiation condition and 35% in the non-differentiation condition. The differentiation effect—the difference in modification rate between the differentiation and the non-differentiation conditions—was larger in the context-cued condition than the non-cued condition (16% vs. 7%).

We analyzed the modification rate data in a logistic mixed-effects model with differentiation condition (non-differentiation vs. differentiation) and contextual cuing of the discourse history (context-cued vs. non-cued) as mean-centered fixed effects and subjects and items as random intercepts; random slopes were included when possible (see Table 1a in the Supplementary Materials for full model details). A significant main effect of differentiation replicated previous findings, with more modification in the differentiation than the non-differentiation condition ($z = –3.63$, $p < .001$). A marginal interaction between differentiation and contextual cuing ($z = 1.76$, $p = .08$) was due to a larger differentiation effect when contexts were cued ($z = –3.51$, $p < .001$) than when they were not ($z = –1.30$, $p = .19$).

3.3.2. Recognition memory

To visualize the participants' ability to discriminate between old and new items, Fig. 3 plots performance in terms of *d'*, a standard index of recognition memory based on signal-detection theory. Inferential statistics for the recognition memory data were evaluated separately for target, contrast, and context items in three planned logistic mixed-effects models. The

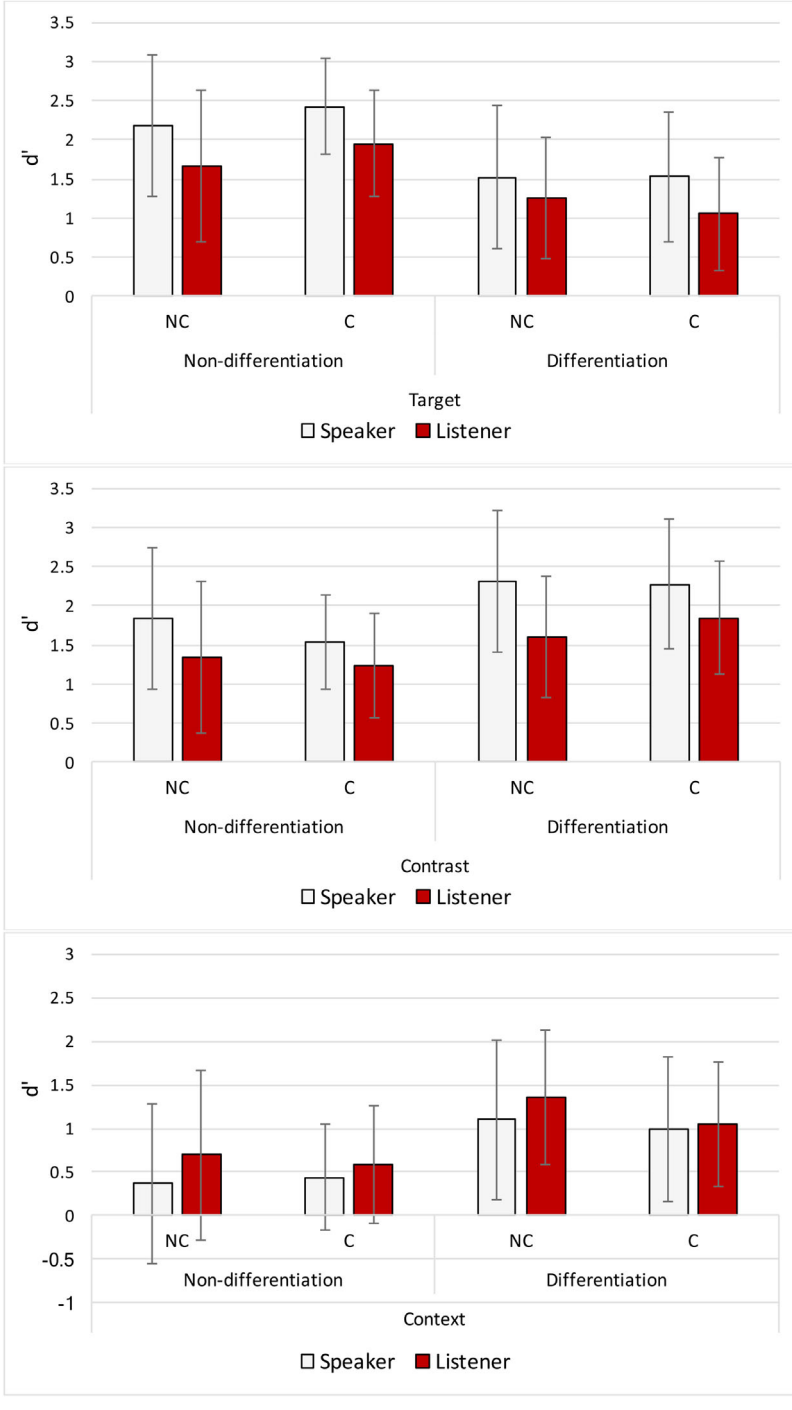

Fig. 3. Experiment 1: Discriminability (*d'*) on the memory test. Error bars indicate by-participant standard deviation (NC: non-cued, C: context-cued).

structure of each model is as follows: The dependent measure is binary: whether the participant responded "yes" ("old" = 1) or "no" ("new" = 0). The intercept term in the model measures response bias—whether participants exhibit a general preference to respond *yes* or *no*. The actual item type (old vs. new) is entered as a centered, fixed effect, and evaluates whether participants respond "old" more often when the item is actually old—this is the primary measure of memory (see Fraundorf, Watson, & Benjamin, 2010; Wright, Horry, & Skagerberg, 2009). Role (speaker vs. listener), differentiation type (non-differentiation vs. differentiation), and contextual cuing (context-cued vs. non-cued) are included as mean-centered fixed effects (see Supplementary Materials B for full model details, Supplementary Tables 1b–3b). The beta weights associated with these fixed effects reflect the effect of these variables on response bias. Of more interest to the present research are the interaction between role and item type (old vs. new) and the interaction between contextual cuing and item type (old vs. new)—in other words, whether the role and contextual cuing affect memory for past items.

For target items (Supplementary Table 1b), speakers had a better memory than listeners ($z = 3.79$, $p < .001$). Both speakers and listeners remembered target items better in the non-differentiation condition than in the differentiation condition ($z = –6.85$, $p < .001$), likely due to the fact that in the non-differentiation condition, the target was the only item from its basic-level object category seen during the communication task. The contextual cue manipulation did not significantly affect memory for targets ($z = –0.98$, $p > .33$).

For contrast items (Supplementary Table 2b), speakers again had a better memory than listeners ($z = 4.58$, $p < .001$). Both speakers and listeners remembered contrast items better in the differentiation condition than in the non-differentiation condition ($z = 2.70$, $p = .007$). The contextual cue manipulation did not significantly affect memory for contrast items ($z = –0.56$, $p > .58$).

For context items (Supplementary Table 3b), *listeners* had a better memory than speakers ($z = –2.15$, $p = .03$), and both speakers and listeners remembered context items better in the dfferentiation versus non-differentiation condition ($z = 3.94$, $p < .001$). The contextual cue manipulation did not significantly affect memory for context items ($z = 0.59$, $p = .56$).

3.3.3. Language–memory interactions

A prerequisite of lexical differentiation is that speakers remember the previously experienced contrast item, so that they can distinguish the current referent from it. We examined whether a speaker's recognition memory for the contrast item in the differentiation condition predicts their use of modification when describing the associated target item during the communication task. A mixed-effects model included the binary measure of memory for the contrast item during the memory test (i.e., whether the speaker correctly recognized the contrast item) as a fixed effect. The dependent measure was whether the target was referenced with a modifier or not. The model revealed no significant effect of contrast memory on target modification ($z = –0.70$, $p = .48$), consistent with previous findings (Yoon et al., 2016).

A separate question is whether using a modifier to describe the target modulates memory for that target at the test—for speakers, listeners, or both. A mixed-effects model included role (speaker vs. listener) and whether the target was modified or not as fixed effects. The dependent measure was whether participants recognized the target item during the memory

test. The model revealed significant main effects of modification ($z = 4.38$, $p < .001$) and role ($z = 2.60$, $p = .009$), but no interaction between modification and role ($z = –0.26$, $p = .80$). This finding indicates that the speaker memory for target items is better than the listener memory and that both speakers and listeners are more likely to correctly recognize the target when it had previously been referred to using a modified noun phrase rather than a bare noun phrase.

3.4. Summary

These findings tentatively suggest that it is perceived relevance, not memory limitations, that influences whether speakers design language with respect to the historical discourse context. We observed that the differentiation rate was marginally higher when contextual cues linked the critical trial with past trials ($p = .08$), compared to when the context was not cued. Speakers in the current study produced modifiers at a higher rate (∼48%), compared to previous studies that used similar stimuli (∼25% in Yoon et al., 2016). The relatively high modification rate in the current study may owe to the use of more complex scenes and the grouping of pictures into sets of 15. Although Yoon et al. (2016) also presented four pictures at a time, the pictures were not grouped across trials using contextual cues. In Experiment 2, we use a more robust manipulation of the discourse context and a within-subjects design to provide a stronger test of the hypothesis that physical cues that highlight the relevance of the discourse history drive speakers to design referring expressions with respect to the historical discourse context.

Analyses of memory for the discourse history also demonstrated a speaker benefit for past referents (target and contrast items), consistent with prior studies of memory for conversation (McKinley, Brown-Schmidt, & Benjamin, 2017; Yoon et al., 2016). The fact that listeners outperformed speakers in context memory (the viewed but unnamed context items on setup trials) is consistent with some prior work in less naturalistic conversational settings (Fischer, Schult, & Steffens, 2015). Although speakers clearly designed their referring expressions with respect to the discourse history, we found that variability in recognition memory for past referents did not predict lexical differentiation, consistent with our earlier work (Yoon et al., 2016). This finding suggests that it is not the strength of the memory trace for the past referent (the contrast item) that determines whether speakers will lexically differentiate but instead, perhaps, whether they perceive that past referent as relevant to the current discourse context.

4. Experiment 2

The aim of Experiment 2 was to replicate and extend the findings of Experiment 1. Changes to the experimental design are as follows: (1) We use a within-subjects manipulation of the contextual cue to increase power. (2) In the context-cued condition, scenic photographs, rather than solid colors, are used as background images to emphasize the distinction between sets. Prior work in the memory literature suggests that a background scene that frames a focal picture can later act as a cue to facilitate remembering *other* pictures that also appeared with that same background scene (Tullis et al., 2014). Thus, in Experiment 2, we use salient

background scenes to perceptually group a series of referents together, with the aim of shaping perceptions about the discourse context.

In the task-based dialogs examined here, the discourse context is strongly determined by the visual stimuli that are presented on each trial. By using detailed and distinct background images to group those stimuli into sets, we hypothesized that we would observe a clear effect of this contextual cue manipulation on speakers' sensitivity to the discourse history.

4.1. Method

4.1.1. Participants

Seventy-two undergraduates (36 pairs) at the University of Illinois at Urbana-Champaign participated in return for partial course credit or cash ($10). Participants were native speakers of North American English. None had participated in Experiment 1. The sample size was based on an a priori power analysis in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007), using the size of the differentiation effect in Experiment 1, calculated using the by-subject condition means for the modification rate in the differentiation condition versus the non-differentiation condition ($d = 0.396$). The power analysis revealed that to achieve 95% power at an alpha level of 5% (one-tailed), the appropriate sample size was 71; data from 72 participants were collected as they participated in pairs.

4.1.2. Materials and procedure

The procedure of Experiment 2 was identical to Experiment 1 with the following exceptions: First, unlike Experiment 1, contextual cuing was manipulated in a blocked, within-subjects design. Across a total of eight blocks, participants alternated between the two context conditions (whether the context was cued, or not, was manipulated across blocks) and, orthogonally, between the roles of speaker and listener. Second, the backgrounds in the context-cued condition were scenic photographs selected to make each set more distinctive (Fig. 4). In the non-cued condition, the background was gray for all sets, just as in Experiment 1.

4.2. Results

4.2.1. Referential communication task

We coded the speakers' referential expressions in the same way as in Experiment 1 (Table 1). The use of modifiers on test trials was analyzed in a logistic mixed-effects model (see Table 2a in the Supplementary Materials). Consistent with the results of Experiment 1, we observed a robust differentiation effect (non-cued: 10% vs. context-cued: 22%; $z = –5.15$, $p < .001$). Further, the interaction between differentiation type and contextual cuing was significant ($z = 2.01$, $p = .04$) due to a larger differentiation effect in the context-cued condition than the non-cued condition.

4.2.2. Recognition memory

The results of the recognition memory test (Fig. 1b) were similar to Experiment 1 and are summarized in Tables 4b–6b in the Supplementary Materials. While past referents were remembered better by speakers than listeners (targets: $z = 3.76$, $p < .001$; contrasts: $z = 4.81$,

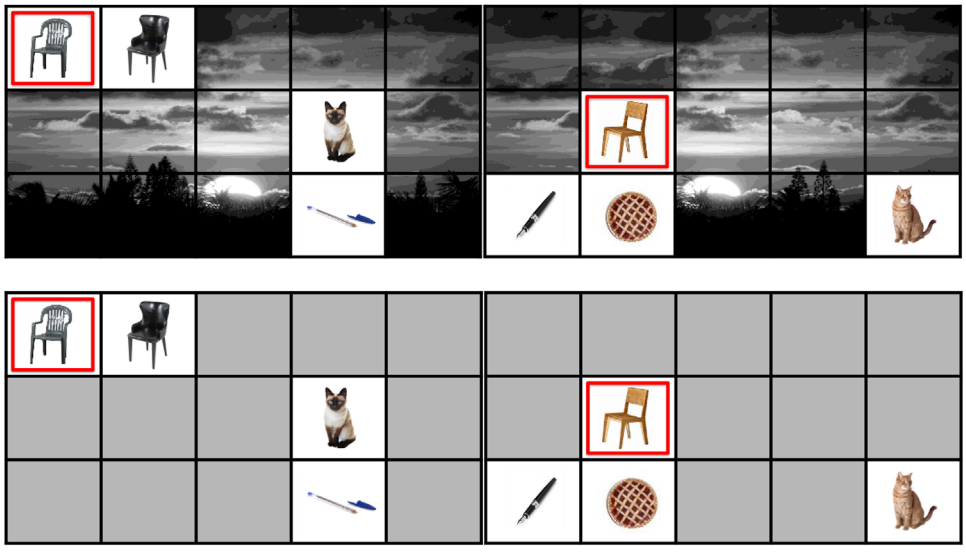

Fig. 4. Experiment 2: Example stimuli during the referential communication task, including setup trials in the differentiation condition (left) and test trials (right), in the context-cued condition (top) and the noncued conditions (bottom). Setup trials in the non-differentiation condition would feature unrelated items (e.g., ball and chimpanzee) instead of chairs.

$p < .001$), this speaking benefit did not extend to the context items that were seen but not named ($z = 0.33$, $p = .74$). Participants exhibited better memory for target items in the non-differentiation than the differentiation condition ($z = –3.31$, $p < .001$), whereas they remembered more contrast items in the differentiation than in the non-differentiation condition ($z = 2.04$, $p = .04$), replicating the results of Experiment 1. There was no evidence that the contextual cuing manipulation influenced interlocutors' memory for any of the past images (target: $z = –0.49$, $p = .63$; contrast: $z = –0.26$, $p = .79$; context: $z = 0.90$, $p = .37$). For context items only, a significant three-way interaction between item type, role, and differentiation ($z = 1.97$, $p = .049$) was obtained. Follow-up analyses were conducted separately for the differentiation and non-differentiation conditions to explore the interaction: In the differentiation condition, there was a marginally significant effect of role ($z = 1.77$, $p = .076$) such that speakers remembered context items better than listeners. In the non-differentiation condition, speaker and listener memory for context items did not significantly differ ($z = –0.65$, $p = .52$).

Last, in analyses relating to language and memory, replicating prior findings, speakers' memory for the contrast object in the differentiation condition did not predict the use of modifiers on test trials ($z = –0.15$, $p = .88$). The target pictures that had been described with a modifier were remembered marginally better than targets that were not described with a modifier ($z = 1.71$, $p = .087$); the interaction between target modification and the role was also not significant ($z = –0.58$, $p = .56$), suggesting that, if modification boosted target memory, there was no evidence that it did so differentially for speakers and listeners.

4.3. Summary

The larger differentiation effect in the context-cued condition is consistent with our hypothesis that perceptions about what constitutes the relevant discourse context shape the design of referring expressions. While the contextual cue manipulation successfully increased the differentiation rate, it had no influence on memory (see Fig. 1b and Supplementary Tables 4b–6b). Clearly, speakers must remember the discourse history in order to be sensitive to it when designing referential expressions. However, the fact that measures of memory were not predictive of the differentiation effect suggests that remembering the discourse history is not sufficient to prompt differentiation. Instead, these findings indicate that perception of the past discourse as *relevant* is a critical component in whether speakers design language with respect to the discourse history.

5. Experiment 3

By design, the contextual cue manipulation in Experiments 1 and 2 involved the simultaneous manipulation of two factors hypothesized to be relevant to the considerations of whether the past discourse history is relevant or not: (1) the use of *background* pictures that grouped the items in a given set of pictures into the same context, and (2) the stability of each picture's location within the scene. These manipulations are analogous to different aspects of the contexts that frame the everyday conversation. For example, the background images are analogous to the physical spaces that frame where a conversation is held (e.g., office, street, etc.), whereas the placement of pictures in each scene is comparable to the placement of objects in a space (e.g., placement of chairs in an office or cars on a street). In Experiments 1–2, we covaried these factors in order to magnify the impact of this manipulation of contextual cue. In Experiment 3, we parametrically manipulated these two factors to examine their independent contributions to the perception of discourse context. While Experiments 1–2 show that the simultaneous manipulation of backgrounds and locations induces the perception of a discourse context linking trials within a set, whether one of these factors alone could drive the effect is unknown. In particular, the background pictures may be sufficient to drive the effect as the change in background image from set to set was highly salient, and it is known that background images of this type can produce reminding effects, shaping language use (Tullis et al., 2014).

5.1. Method

5.1.1. Participants

One hundred forty-four undergraduates (72 pairs) at the University of Illinois at Urbana-Champaign participated in return for partial course credit or cash ($10). The minimum sample size ($N = 71$) was based on the same a priori power analysis as in Experiment 2; since participant role (speaker vs. listener) was manipulated between participants, data from 72 pairs were collected. Participants were native speakers of North American English. None had participated in Experiments 1 and 2.

5.1.2. Materials and procedure

The procedure of Experiment 3 was similar to Experiment 2: A referential communication task followed by a memory test. Unlike Experiment 2, participant role (speaker vs. listener) was manipulated between participants. The two participants were assigned to the roles of speaker and listener and maintained that role throughout the experiment.

In the referential communication task, two variables, picture position (fixed vs. random) and background (gray vs. unique background), were manipulated in a blocked, within-subjects design. Thus, together with the differentiation manipulation (differentiation vs. non-differentiation), the experiment was a 2 × 2 × 2 within-participants design. In order to rotate test items across the eight conditions created by the combination of picture position (fixed vs. random), background (unique vs. gray), and differentiation (differentiation vs. non-differentiation), we created eight lists and reduced the number of target items to 24. Each participant was exposed to 24 different sets of items (e.g., boots, dogs, socks). As in Experiments 1 and 2, each item set in the referential communication task was associated with one setup trial, one test trial, and six filler trials (totaling 24 setup, 24 test, and 144 filler trials). The memory test was old-new recognition and consisted of 144 trials total (24 old and 24 new items for each target, contrast, and context items, respectively).

5.1.3. Predictions

We expected to replicate the previous findings in the fixed/unique background condition and the random/gray background condition. These two conditions are the same as those used in Experiment 2 and that showed that the discourse history shapes referential form to a greater extent when contextual cues group the present and past into the same discourse context. The primary question that Experiment 3 is poised to address is whether the observed effect of the contextual cue manipulation on lexical differentiation was due to the use of fixed picture positions across trials, the background cue, or both.

5.2. Results

5.2.1. Referential communication task

The differentiation effect (Table 1) was largest in the unique/fixed condition (17%) and smallest in the gray/random condition (9%), with the unique/random (16%) and gray/fixed condition (11%) in between. The data were analyzed with a logistic mixed-effects model (see Table 3a in the Supplementary Materials). Although speakers produced significantly more modifiers in the differentiation condition than the non-differentiation condition ($z = –4.09$, $p < .001$), neither the effect of background ($z = –0.1$, $p = .92$) nor the effect of picture position ($z = 0.003$, $p = 1.00$) was significant. The interaction between background and picture position was not significant ($z = 0.25$, $p = .80$). The interaction between differentiation (non-differentiation vs. differentiation) and background ($z = 0.84$, $p = .40$) and the interaction between differentiation and picture position ($z = 1.07$, $p = .28$) were also not significant. We note, however, that the effects for the two conditions that were identical to Experiment 2—the fixed/unique background and the random/gray background conditions—were in the same direction as the results of Experiment 2.

5.2.2. Recognition memory

For target items (Fig. 2b, Supplementary Table 7b), speakers outperformed listeners (z = 4.90, $p < .001$). Participants remembered target items better in the non-differentiation than in the differentiation condition, likely due to the fact that only a single item from that category was seen during the conversation, minimizing within-category interference (z = -5.92, $p <.001$). These effects were qualified by a five-way interaction between item, role, differentiation, background, and picture position ($z = 2.82$, $p = .005$). Exploration of the interaction indicated that a four-way interaction between item, role, differentiation, and background was significant in the fixed position condition ($z = –4.22$, $p < .001$) but not in the random position condition ($z = 0.65$, $p = .52$). The significant four-way interaction between item, role, differentiation, and background in the fixed position condition was driven by a significant three-way interaction between item, differentiation, and role in the fixed-gray condition (z = –3.50, $p < .001$) but not in the fixed-unique condition (1.75, $p = .08$). In the fixed-gray condition, *speakers* remembered target items better in the non-differentiation condition than in the differentiation condition ($z = –2.89$, $p = .004$), whereas listeners' memory did not differ between the two conditions ($z = –0.47$, $p = .64$).

For contrast items (Supplementary Table 8b), speakers outperformed listeners ($z = 4.58$, $p < .001$), and the memory was better in the differentiation condition ($z = 2.49$, $p = .01$). These effects were qualified by a significant four-way interaction between item, role, differentiation, and background ($z = 2.10$, $p = .04$). Exploration of this interaction indicated that the three-way interaction between item, role, and differentiation was significant in the unique condition ($z = –2.92$, $p = .004$) but not in the gray condition ($z = 0.14$, $p = .89$). The three-way interaction between item, role, and differentiation in the unique condition was driven by the fact that speakers' memory between the non-differentiation and the differentiation conditions did not differ ($z = –0.53$, $p = .60$), whereas listeners' memory was non-significantly better in the differentiation condition than in the non-differentiation condition ($z = 1.64$, $p = .10$).

For context items (Supplementary Table 9b), there was no speaking benefit ($z = –0.15$, $p = .88$); however, the memory was better in the differentiation condition ($z = 5.64$, $p < .001$).

Last, analyses relating memory and language again demonstrated that the speaker's memory for contrast items in the differentiation condition did not predict their use of modifiers when describing the associated target ($z = –1.08$, $p = .28$). However, we did find that for both speakers and listeners, when the speaker chose to modify the target, memory for the target improved at test ($z = 2.58$, $p = .01$). Speakers' memory for target was better than listeners' memory ($z = 3.93$, $p < .001$), but the interaction between the use of modifiers and role was not significant ($z = –0.56$, $p = .57$).

5.3. Summary

The results of Experiment 3 demonstrate a lexical differentiation effect, but the independent manipulations of background context and picture position failed to reach significance. The pattern of effects, however, was consistent with previous findings.

Consistent with the results of Experiments 1 and 2, speakers remembered mentioned referents better than listeners, but the memory for past contexts did not predict lexical differentiation.

6. Experiment 4

The aim of Experiment 4 was to directly replicate Experiment 2. While the parametric manipulations of contextual cues in Experiment 3 did not lead to statistically significant condition differences, the study may have suffered from a lack of power because of partitioning the background and picture position variables into four separate conditions. Also, consider that in everyday life, physical contexts (e.g., offices) may systematically cue the location of objects within them (e.g., chairs). If so, the independent manipulation of background and picture position in Experiment 3 may have caused participants to ignore these grouping cues. Thus, in running a direct replication of Experiment 2, Experiment 4 is positioned to demonstrate the reproducibility of the contextual cuing effect. If the combination of a background cue and stable picture positions together increases the perception of a discourse context that links trials within a set, the differentiation effect should be larger in the context-cued compared to the non-cued condition.

6.1. Method

6.1.1. Participants

One hundred and four undergraduates (52 pairs) at the University of Illinois at Urbana-Champaign participated in return for partial course credit or cash ($10). Participants were native speakers of North American English. None had participated in Experiments 1–3. An a priori power analysis based on the size of the differentiation effect in Experiment 1 ($d = 0.396$) indicated that to achieve the power of 99% at an alpha level of 5% (one-tailed), the minimum number of participants was 102. While Experiment 4 is a direct replication of Experiment 2, we based this power analysis on the size of the differentiation effect in Experiment 1 because the effect size was smaller in Experiment 1 than it was in Experiment 2 (E1: $d = 0.396$; E2: $d = 0.534$), and we wished to be conservative in estimating the magnitude of the differentiation effect.

6.1.2. Materials and procedure

The procedure of Experiment 4 was identical to Experiment 2.

6.2. Results

6.2.1. Referential communication task

The differentiation effect was 17% in the context-cued condition versus 7% in the non-cued condition (Table 1), a significant effect ($z = –4.82$, $p < .001$; Supplementary Table 4a). A significant interaction between differentiation and the contextual cue manipulation ($z = 2.09$, $p = .04$) was due to a larger differentiation effect in the context-cued condition than the non-cued condition. These findings replicate the results of Experiment 2.

6.2.2. Recognition memory

Performance on the memory task was also consistent with the results of Experiment 2 (Fig. 3b and Supplementary Tables 10b–12b). Speakers were better than listeners at remembering target ($z = 4.48, p < .001$) and contrast items ($z = 6.81, p < .001$) but did not remember more context items ($z = –0.90$, $p = .37$). The significant main effects of differentiation for target, contrast, and context items were also replicated: Participants remembered target items better in the non-differentiation than in the differentiation condition ($z = –5.14$, $p < .001$), while memory for contrast items ($z = 2.45$, $p = .01$) and context items ($z = 5.44$, $p < .001$) was higher in the differentiation condition. In all three models, the discourse context effect was not significant (target: $z = 1.31$, $p = .19$; contrast: $z = 1.76$, $p = .08$; context: $z = 0.70$, $p = .49$). For target items, a significant three-way interaction between item, role, and differentiation ($z = –2.16$, $p = .03$) was driven by the fact that the effect of differentiation on memory was larger for speakers ($z = –4.42$, $p < .001$) than for listeners ($z = –3.21$, $p = .01$). For contrast items, a significant three-way interaction between item, differentiation, and context ($z = 3.74, p < .001$) was driven by the fact that in the non-cued condition, participants remembered contrast items better in the differentiation condition than in the non-differentiation condition ($z = 6.20$, $p < .001$), whereas performance did not differ between the two conditions when the context was cued ($z = 0.89$, $p = .37$).

Finally, speakers' memory for the contrast item did not predict modification of the target during the communication task ($z = 0.50$, $p = .62$). Again, speakers remembered the target better than listeners ($z = 4.47, p < .001$), and both speakers and listeners remembered target items better when the speaker used a modifier ($z = 4.18$, $p < .001$). The interaction between the use of modifiers and role was not significant ($z = –0.64$, $p = .52$).

6.3. Summary

The results of Experiment 4 replicate the findings of Experiment 2. We observed a clear differentiation effect that was magnified by the contextual cue manipulation, where trials within a set were perceptually linked using the same background scene and fixed picture positions. Again, we found that while speakers exhibit better memory for past referents than listeners, this speaking benefit did not extend to unmentioned items in the discourse context. Memory was also not influenced by the contextual cue manipulation.

7. Joint analysis of Experiments 1–4

The aim of this analysis was to combine the data across the four studies to have a more powerful analysis of the key effects. Several of the critical effects reported in this manuscript were significant at levels that, in and of themselves, might evoke concern over replicability (cf. Patil, Peng, & Leek, 2016). A joint analysis enables more precise testing of the central hypotheses and estimation of the relevant effect sizes.

7.1. The differentiation effect

The referential data from Experiments 1 to 4 were combined, excluding the data from the gray/fixed and unique/random conditions in Experiment 3, which could not be easily

Table 2
Proportion of modified expressions (SD) at test trials during the referential communication task across Experiments 1–4

Contextual Cue	*Non-differentiation*	*Differentiation*
Non-cued	35.5% (30.0), $N = 284$	43.8% (31.1), $N = 284$
Context-cued	31.6% (29.3), $N = 284$	49.6% (31.9), $N = 284$

integrated with the other datasets (Table 2). Across the four datasets, the lexical differentiation effect was about twice as large in the context-cued condition (18%) than in the non-cued condition (8%). A combined statistical analysis revealed a significant interaction between differentiation and contextual cuing ($z = 3.71$, $p < .001$; Supplementary Table 5a) driven by a larger differentiation effect in the context-cued ($z = –8.59$, $p < .001$) than the non-cued condition ($z = –3.97$, $p < .001$). This result provides clear evidence that speakers are more likely to take the discourse history into consideration when contextual cues create the perception of a discourse context that links current and past referents.

7.2. Memory and language use

The data from the differentiation conditions in Experiments 1–4 were combined to provide a stronger test of whether remembering the contrast predicts modification during the communication task. Consistent with the analyses from the individual experiments, the speaker's memory *did not predict* their use of modification ($z = 0.82$, $p = .41$).

7.3. Memory for contexts in conversation

The memory data for the context items from the setup trials in experiments 1–4 were combined (Fig. 5) and examined as a function of role (speaking vs. listening) and differentiation (non-differentiation vs. differentiation) to explore how interlocutors represent unmentioned context items. We examined context items in the differentiation and non-differentiation conditions together, although the specific items differed across the two conditions: Recall that in the differentiation condition, the contrast and context items were related (e.g., two chairs), whereas in the non-differentiation condition, the contrast and context items were unrelated (e.g., ball and chimpanzee).

Memory for context items was analyzed in a logistic mixed-effect model. In the model, role (speaker vs. listener), differentiation (non-differentiation vs. differentiation), and item type (old vs. new) were included as fixed effects (see Supplementary Materials, Table 13b). The interaction of role and differentiation with the item type variable are used to estimate the effect of these variables on memory. Both speakers and listeners remembered context items better when they were related to the named contrast item (differentiation condition) than when they were unrelated to the named contrast item (non-differentiation condition; $z = 6.33$, $p < .001$). Speakers' and listeners' memory for context items did not differ ($z = –1.25$, $p = .21$), and the interaction between role and differentiation was not significant ($z = 1.41$, $p = .16$).

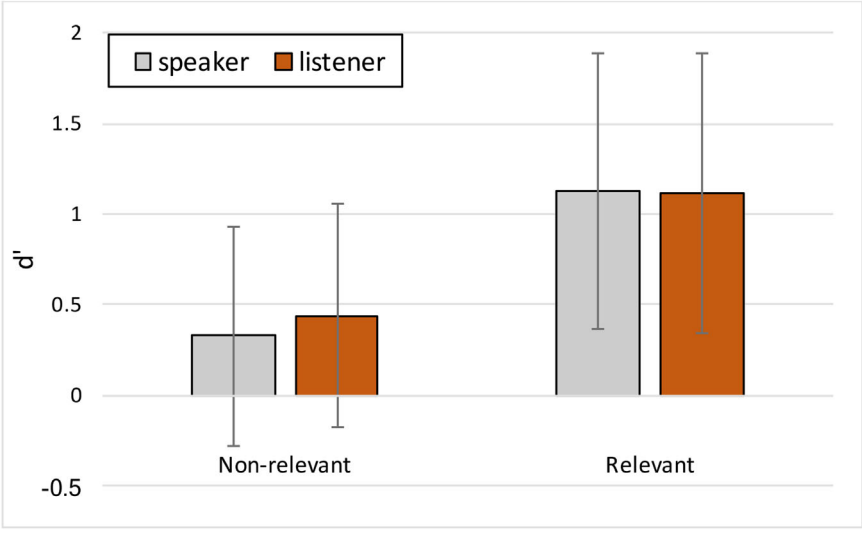

Fig. 5. Joint analysis of Experiments 1–4: Discriminability (*d'*) on the recognition test for context items in the non-differentiation condition (non-relevant; e.g., memory for chimpanzee when the named object is a ball) and the differentiation condition (relevant; e.g., memory for a leather chair when the named object is a plastic chair). Error bars indicate the standard deviation of the by-participant means.

Taken together, the speaking benefit in memory that was observed for mentioned items did not extend to unmentioned items; speakers and listeners remembered unmentioned items similarly. In addition, we observed that speakers and listeners remembered relevant context items better than non-relevant items.

8. General discussion

In four experiments, participants engaged in task-based conversation in which they named a series of objects in complex visual displays. We manipulated properties of those visual displays in order to shape perceptions about the relevant discourse context, linking current with past referents. Using the lexical differentiation effect (Van Der Wege, 2009) as a test case, we tested the hypothesis that speakers would exhibit more sensitivity to the discourse history when contextual cues linked the immediate discourse context with past contexts. Consistent with this hypothesis, speakers were more likely to lexically differentiate current from past referents when perceptual cues grouped the current and past referents into the same context. In a joint analysis of Experiments 1–4, we found that the differentiation effect grew to 18% in the context-cued condition from 8% in the non-cued condition. The estimates from the non-cued condition are consistent with the relatively low differentiation rates observed in prior work (∼8%) in which speakers described referents over a series of unrelated trials (Yoon & Brown-Schmidt, 2013; Yoon et al., 2016). The present findings point to a clear role for contextual cues in shaping perceptions about whether the discourse history is relevant to current

discourse. Entities considered to be a part of the current discourse context are part of the relevant context that shapes referential form and perceptual boundaries define the limits of that influence.

An important open question regarding the nature of the influence of discourse context on referential design is what specific factors determine the speaker's perception of the discourse context over time. In Experiments 2 and 4, we successfully cued the perception of a discourse context through the use of stable picture positions and distinct photographic background images. While our attempts to parametrically manipulate these two variables in Experiment 3 did not lead to the successful demonstration of a contextual cuing effect, inspection of the means in Experiment 3 suggests that the two factors may have roughly additive effects. The lack of significant contextual cuing effects in Experiment 3 may owe to a lack of power. Also possible is that the parametric manipulation of the background cue and picture positions in Experiment 3 disrupted typically occurring links between these cues in everyday life, causing participants to rely less heavily on them. By contrast, in Experiment 4, which directly replicated Experiment 2, these two factors—background and picture locations—covaried in a way that may mimic everyday life and may have contributed to the significant and reproducible effect that was observed.

The literature additionally suggests that factors such as actions and intentions (Grosz & Sidner, 1986), physical constraints such as distance between referents (Beun & Cremers, 1998; Brown-Schmidt & Tanenhaus, 2008; Greene, Gerrig, McKoon, & Ratcliff, 1994), grammatical and event structure (Kaiser, Runner, Sussman, & Tanenhaus, 2009; Kehler & Rohde, 2013), and discourse markers (Byron & Heeman, 1997; Schourup, 1999) all have the potential to shape the perception of whether or not past referents are relevant to the current talk. Whether the discourse context is cued by task-relevant perceptual features, linguistic discourse markers, events, and so on will likely depend on the nature of the conversation itself (e.g., whether participants are talking about visually copresent or abstract entities).

In earlier work, we tested the hypothesis that speakers often fail to remember the discourse history in an attempt to explain why speakers so rarely designed their referring expressions with respect to the historical discourse context (Yoon et al., 2016). However, the present findings, along with our earlier work, conclusively show that memory for past referents is not sufficient to prompt speakers to lexically differentiate: Memory for past referents was good (on average 70.9% correct recognition of targets across Experiments 1–4, compared to a false alarm rate of 13.1%), and it did not predict the differentiation rate. While memory for the discourse history is clearly necessary in order to design language with respect to it, our findings show that the strength of these memory representations is alone not predictive of whether a speaker will design language with respect to the past. That said, in all four experiments, we did observe clear language–memory links, with both speakers and listeners more likely to correctly recognize the target referent if it had previously been described using a modifier (this effect was significant in Experiments 1, 3, 4, and marginal in Experiment 2).

Note that our memory tests are designed to measure the mnemonic strength of items relative to one another after a delay. Although these relative differences in memory are likely to reflect differences in initial learning rather than differential forgetting across the course of the experiment, an open question for future research is whether a measure of mnemonic strength

for the contrast item at the time the speaker prepares to name the target would predict lexical differentiation. An alternative, non-mutually exclusive interpretation of our findings, then, is that as speakers prepared to reference the target, they were more likely to spontaneously retrieve the contrast item from memory, and it is this spontaneous memory retrieval that prompted lexical differentiation. While the fact that contrast memory, measured at a delay, did not predict modification is somewhat inconsistent with this explanation, it remains a candidate mechanism to explore in future work.

In addition, we replicated the finding that speakers remember past referents better than listeners (McKinley et al., 2017; Paterson & Kemp, 2006; Yoon et al., 2016), consistent with the generation effect in the memory literature (Mulligan & Lozito, 2004; Slamecka & Graf, 1978). These findings are consistent with observations that following conversation, people recall more of what they said, compared to what was said to them (Ross & Sicoly, 1979), and are more likely to re-use phrases that they themselves introduced into the discourse (Knutsen & Le Bigot, 2014). Speakers may invest more effort to generate referring expressions and, in doing so, encode the referents better than listeners. Also consistent with our earlier research, we found that this speaking benefit did not extend to undiscussed elements of the discourse context, suggesting that the memorial record of the *context* of the conversation is more similar for speakers and listeners than the record of what was *said* (also see McKinley et al., 2017).

The fact that conversational partners walk away from conversations with *distinct memories for what was discussed* has been described as an egocentric bias (Ross & Sicoly, 1979) and points to limits on the degree to which partners can become aligned in conversation (Pickering & Garrod, 2004). This observation raises questions about how a rational language production system (Frank & Goodman, 2012) should operate given differences in what is remembered about past communicative exchanges (Brown-Schmidt & Heller, 2018). The fact that this speaking benefit is eliminated for elements of the context may hold explanatory power in understanding the loci of communication failures. Alternatively, differences in what is remembered between speakers and listeners may reveal differences in those roles in fostering successful communication. Perhaps speakers process target items with more focal attention, whereas listeners distribute attention more broadly because comprehension is a more error-prone process than production. This explanation is broadly consistent with previous findings of asymmetry between production and comprehension in that speakers, but not listeners, demonstrate sensitivity to the historical context (Yoon & Brown-Schmidt, 2013). More broadly, this explanation of the speaker-listener asymmetry is consistent with the notion of transactive memory (Hirst & Echterhoff, 2012; Hollingshead & Brandon, 2003), wherein groups of people take a divide-and-conquer approach to memorial tasks such that their collaborative remembering can outperform each individual's memory. This idea is similar to classic accounts of conversational language use which posit that conversational partners work together to minimize joint effort (Clark & Wilkes-Gibbs, 1986). Here, we propose that speakers and listeners distribute their labor and, in doing so, give rise to a more effective communicative system than if they perfectly aligned their attention within the conversation. An open question is whether speakers and listeners are aware of the fact that they have asymmetric memory representations and, if so, whether they modulate their language processing accordingly.

9. Conclusion

Considerable evidence supports the view that reference is produced and understood with respect to a context that includes entities both in the immediate physical context and relevant entities from the discourse history (Brennan & Clark, 1996; Gordon et al., 1993; Marslen-Wilson et al., 1982; Poesio et al., 2002). Yet the set of potential referents in the immediate and recent past is more or less unbounded (Gleitman & Gleitman, 1992), and thus requiring the relevant referential domain to be constrained in some way (Brown-Schmidt & Tanenhaus, 2008; Brown-Schmidt, Yoon, & Ryskin, 2015; Chambers et al., 2002; Landragin, 2006). Here, we tested the hypothesis that past referents are considered to be relevant to current discourse when contextual cues point to past referents being part of the same discourse context. We find that when the past is perceptually linked to the present, speakers are more likely to take the past into account when designing referring expressions. We propose that the mechanism by which speakers determine the discourse context is ongoing assessments of relevance, which in turn is shaped by cognitive processes such as perceptions of similarity. We also report a robust replication of prior findings of a speaking benefit in memory for past referents that does not extend to the broader discourse context. This asymmetry in memory for the discourse history places limits on the degree to which coordinated representations of what has been said are possible and offers explanatory power for understanding miscommunication in conversation.

Acknowledgments

This study is based on work supported by National Science Foundation Grant BCS 15–56700 to Sarah Brown-Schmidt.

References

Altmann, G. T. M., & Steedman, M. J. (1998). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The "blank screen paradigmdä. *Cognition*, *93*(2), B79–B87.

Beun, R., & Cremers, A. H. (1998). Object reference in a shared domain of conversation. *Pragmatics & Cognition*, *6*, 121–152.

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., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. *Psychology of Learning and Motivation*, *53*, 301–345.

Brown, G. D., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576.

Brown-Schmidt, S., & Heller, D. (2018). Perspective-taking during conversation. In G. Gaskell & S. A. Rueschemeyer (Eds.), *Oxford handbook of psycholinguistics*. (pp. 551–574). Oxford, England: Oxford University Press.

Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, *32*, 643–684.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as Contexts in Conversation. *Psychology of Learning and Motivation*, *62*, 59–99.

Byron, D. K., & Heeman, P. A. (1997). Discourse marker use in task-oriented spoken dialog. In G. Kokkinakis, N. Fakotakis & E. Dermatas (Eds.), Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech '97) (pp. 2223–2226). Patras, Greece: University of Patras, Wire Communications Lab.

Carpenter, S. K., Pashler, H. P., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, *36*, 438–448.

Chambers, C. G., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. *Cognition*, *108*(1), 26–50.

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.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 687–696.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review*, *123*, 324–347.

Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Theoretical issues in natural language processing* (pp. 57–63). New York: Association for Computing Machinery.

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

Elmes, D. G., Adams, C., & Roediger, H. L. (1970). Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology*, *86*, 103–107.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

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*, 195–204.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998–998.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language*, *63*, 196–227.

Gleitman, L. R., & Gleitman, H. (1992). A picture is worth a thousand words, but that's the problem: The role of syntax in vocabulary acquisition. *Current Directions in Psychological Science*, *1*, 31–35.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, *17*, 311–347.

Greene, S. B., Gerrig, R. J., McKoon, G., & Ratcliff, R. (1994). Unheralded pronouns and management by common ground. *Journal of Memory and Language*, *33*, 511–526.

Grosz, B. J., & Sider, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*, 175–204.

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.

Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition*, *149*, 104–120.

Hilliard, C., O'Neal, E., Plumert, J., & Cook, S. W. (2015). Mothers modulate their gesture independently of their speech. *Cognition*, *140*, 89–94.

Hirst, W., & Echterhoff, G. (2012). Remembering in conversations: The social sharing and reshaping of memories. *Annual Reviews of Psychology*, *63*, 21.1–21.25.

Hollingshead, A. B., & Brandon, D. P. (2003). Potential benefits of communication in transitive memory systems. *Human Communication Research*, *29*, 607–615.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addresses. *Journal of Memory and Language*, *47*(4), 589–606.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*, 127–142.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541–573.

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.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition*, *112*, 55–80.

Kehler, A., & Rohde, H. (2013). A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics*, *39*, 1–37.

Knutsen, D., & Le Bigot, L. (2014). Capturing egocentric biases in reference reuse during collaborative dialogue. *Psychonomic Bulletin & Review*, *21*(6), 1590–1599.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from tower of Hanoi. *Cognitive Psychology*, *17*, 248–294.

Landragin, F. (2006). Visual perception, language and gesture: A model for their understanding in multimodal dialog systems. *Signal Process*, *86*, 3578–3595.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozbuko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 671–685.

Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. In R. J. Jarvella & W. Klein (Eds.), Speech, Place, and Action. (pp. 339–378). Chichester: Wiley.

McKinley, G. L., Brown-Schmidt, S., & Benjamin, A. S. (2017). Memory for conversation and the development of common ground. *Memory & Cognition*, *45*, 1281–1294.

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. *Psychology of Learning and Motivation*, *45*, 175–214.

Nadig, A. S., & Sedivy, J. C. (2002). Evidence of perspective-taking constraints in children's online reference resolution. *Psychological Science*, *13*, 329–336.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Oxford, England: Prentice-Hall.

Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review*, *77*, 257–273.

Osgood, C. E. (1971). Where do sentences come from? In D. D. Steinberg & L. A. Jakobovits (Eds.), Semantics: An interdisciplinary reader in philosophy, linguistics, and psychology. (pp. 497–529). New York: Cambridge University Press.

Paterson, H. M., & Kemp, R. I. (2006). Comparing methods of encountering post-event information: The power of co-witness suggestion. *Applied Cognitive Psychology*, *20*, 1083–1099.

Patil, P., Peng, R. D., & Leek, J. T. (2016). What should researchers expect when they replicate studies? A statistical view of replicability in psychological science. *Perspectives on Psychological Science*, *11*, 539–544.

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*(2), 169–190.

Poesio, M., Di Eugenio, B., & Keohane, G. (2002). *Discourse structure and anaphora: An empirical study*. NLE Technical Note No. TN-02-02: University of Essex.

Reitman, W., Marlin, J. T., Bjork, R. A., & Higman, B. (1973). Strategy control and directed forgetting. *Journal of Verbal Learning and Verbal Behavior*, *12*, 140–149.

Ross, M., & Sicoly, F.Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*(3), 322–336.

Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, *43*(8), e12769.

Schourup, L. (1999). Discourse markers. *Lingua*, *107*(3-4), 227–265.

Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Leaning and Verbal Behavior*, *6*, 156–163.

Siler, J., & Benjamin, A. S. (2020). Long-term inference and memory following retrieval practice. *Memory & Cognition*, *48*, 645–654.

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

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 384–397.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220.

Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy*, *25*, 701–721.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tetreault, J., & Allen, J. (2003). An empirical evaluation of pronoun resolution and clausal structure. *Proceedings of the 2003 International Symposium on Reference Resolution and its Applications to Question Answering and Summarization* (pp. 1–8). Venice, Italy: Universit Ca Foscari.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergartenpath effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134.

Tullis, J. G., Braverman, M., Ross, B. H., & Benjamin, A. S. (2014). Remindings influence the interpretation of ambiguous stimuli. *Psychonomic Bulletin & Review*, *21*, 107–113.

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. (1998). Centering, anaphora resolution, and discourse structure. In M. A. Walker, A. K. Joshi & E. F. Prince (Eds.), *Centering theory in discourse* (pp. 401–435). Oxford, England: Oxford University Press.

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

Wolter, L., Gorman, K. S., & Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of Memory and Language*, *65*(3), 299–317.

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, *41*, 257–267.

Yoon, S. O. (2018). The role of discourse context in reference production and comprehension: Insights from the lexical differentiation effect. *Psychology of Learning and Motivation*, *68*, 361–392.

Yoon, S. O., Benjamin, A., & Brown-Schmidt, S. (2016). The historical context in conversation: Lexical differentiation and memory for the discourse history. *Cognition*, *154*, 102–117.

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