This paper published as: Ahn, S-O & **Brown-Schmidt**, S. (in press). Retrieval processes and Audience design. *Journal of Memory and Language.*

Retrieval processes and Audience design

So Eun Ahn[1]

Sarah Brown-Schmidt

Vanderbilt University

Corresponding Author:

Sarah Brown-Schmidt

Vanderbilt University Department of Psychology and Human Development

230 Appleton Place · Nashville, TN 37203, USA

615-875-8081

sarahbrownschmidt@gmail.com

[1] Present Address: So Eun Ahn is now at the University of California, San Diego

Abstract

Conversational partners develop shared knowledge. In referential communication tasks, partners collaboratively establish brief labels for hard-to-name images. These image-label mappings are associated in memory with that partner, evidenced by use of those brief labels with the same partner, and longer descriptions with new partners. According to the people-as-contexts view, the conversational partner functions as a contextual cue to support retrieval of conversationally-relevant information. Inspired by findings from the memory literature that context effects can be stronger when retrieval is more explicit, two experiments test the hypothesis that the speaker will be more likely to invoke the partner as a retrieval cue when retrieval processes are more explicit. The results indicated a strong effect of partner that, contrary to these predictions, was not boosted by explicit retrieval processes. The lack of an effect of retrieval processes speaks to the ubiquity with which language use in conversation is tailored to the particular people with whom we converse.

Keywords: audience design, memory, conversation, retrieval processes, partners as contexts

Introduction

In conversation, each person has a distinct perspective. For conversational partners, some aspects of their perspectives are shared and part of their common ground, whereas other aspects of their perspectives are private and part of their privileged ground. Keeping track of what information is shared vs. privileged is thought to be a critical pre-requisite to engaging in a successful conversation. When a speaker refers to something, she must evaluate how familiar the addressee is with the referent, and then select a referring expression that is specific enough for the addressee to identify it (e.g., *he* vs. *the baby vs. the baby in the dinosaur onesie*). This basic aspect of everyday conversation—referring—requires that speakers use information about what specific individuals know and don't know. Addressees have similar burdens; identifying what a speaker is referring to relies on beliefs about the speaker's knowledge.

Audience Design

Current approaches to perspective-taking are greatly influenced by the work of Clark and colleagues who proposed that individuals maintain representations of common ground they share with specific individuals (Clark, 1992, 1996; Clark & Wilkes-Gibbs, 1986; also Stalnaker, 1978). Early evidence that conversational partners develop shared knowledge in conversation came from analyses of the labels for game-pieces that interlocutors establish as they engage in a referential communication task (Krauss & Weinheimer, 1966). The task requires that the partners repeatedly refer to the game pieces over a series of rounds. The game-pieces are typically abstract images that lack conventional names. Thus as the partners refer to the images, initially lengthy game-piece descriptions, e.g., *the one that looks like it could be a person dancing...,* are jointly shaped into brief labels, e.g., *"the ballet dancer*" (Krauss &

Weinheimer, 1964, 1966; Clark & Wilkes-Gibbs, 1986). A key finding is that speakers use these labels in a partner-specific manner. Once the two partners develop shared labels for the game-pieces, if one of the partners is then replaced by a naïve partner, the partner with more experience in the task reverts to longer descriptions when speaking to the new partner (Wilkes-Gibbs & Clark, 1992; Yoon & Brown-Schmidt, 2018; 2019). These longer descriptions tend to be elaborations on labels established with the familiar partner, for example, "*the O one"* might be lengthened to "*the like O with the line across"*. This use of longer descriptions is thought to allow the new partner to succeed in the task, as naïve partners have difficulty interpreting the brief labels established by other partners (Schober & Clark, 1989). The speaker's sensitivity to the knowledge of the conversational partner is known as the process of *audience design* – the use of language in a way that is tailored to the knowledge and perspective of one's addressee (Clark, 1996; Clark & Murphy, 1982; Sacks, Schegloff, & Jefferson, 1978; Ferreira, 2019). A large body of evidence now shows that, to varying degrees in different circumstances, speakers adjust both their speech and gestures based on the knowledge, beliefs, and identity of the addressee or addressees (Wardlow Lane et al., 2006; Brennan & Clark, 1996; Horton & Gerrig, 2002; Horton & Spieler, 2007; Yoon & Brown-Schmidt, 2018; 2014; Hilliard & Cook, 2016; Galati & Brennan, 2010).

Building on this idea that conversational partners use language in a partner-specific manner, the present research examines the processes that guide how these partner-specific representations of perspective are accessed from memory. The idea that partners serve as memory cues was introduced by Horton and Gerrig (2005) as an explanation for how common ground might be represented and accessed in conversation. They argued that partners serve as contextual cues for accessing related information in memory through a resonance process (also see Horton and Gerrig, 2016). For example, Horton and Gerrig (2005) found that speakers were more successful at designing utterances for specific dialogue partners when each of two partners was associated with pictures from different categories, compared to a case where the two partners were associated with pictures from the same categories (see MacLeod, et al., 2010; Hunt & McDaniel, 1993 for relevant discussion). On their model, audience design processes can either be automatic or strategic (see Horton & Gerrig, 2005, 2016; Horton, 2007). While the time-course of audience design is not a focus of the present research, considerations of timing and computational effort are one motivation for the idea that audience design may proceed automatically in some cases (see Horton and Brennan, 2016).

Insights from the Memory Literature

A classic finding in the memory literature is that recall of studied items improves when the study context is reinstated at test, compared to a case where participants are tested in a new context (e.g., Godden & Baddeley, 1975; Smith, 1979). Brown-Schmidt, et al. (2015) argued that much like a physical location might cue retrieval of studied words, the conversational partner that you are addressing in conversation acts as a type of context that can cue retrieval and subsequent use of associated information in conversation. Thinking of partners as contexts allows us to draw on findings in the context memory literature to make predictions about the mechanisms by which partners will influence language use in conversation. For example, in a review and meta-analysis, Smith and Vela (2001) report that incidental environmental context effects are significantly amplified by a change in experimenter, such that the effect of testing in a new room is only d=.26 when the experimenter stays the same from study to test, but is considerably larger, *d*=.62 when the experimenter is also new. Thus partners may be a particularly effective cue to retrieve information from memory, possibly because persons are an integral part of social interactions. Further, some evidence indicates that when participants integrate the studied items with the environmental context, the effect of same vs. new context at test is amplified (Eich, 1985). Extended to conversation, these findings predict that specific discourse partners should be bound in memory to the contents of talk when they are seen as integral to the talk at hand. Take the case of the terms developed in a referential communication task to refer to abstract game pieces (e.g. "the ballet dancer"); in such settings, partners expect one another to continue using the same terms once they have been established (Metzing & Brennan, 2003). Consistent with what we would predict from the findings in the memory literature, partner-specific effects are in fact attenuated in non-interactive settings where the partner is not an integral part of the interaction (Brown-Schmidt, 2009; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; also see Knutsen & LeBigot, 2014). Moreover, if a partner expresses a lack of attention during conversation, the information discussed is not assumed to be part of the common ground (Craycraft & Brown-Schmidt, 2018). Of note is that these partner-specific effects depend on who the speaker intends to address in the communicative setting. In a situation with three people in the room, and the speaker alternates in addressing one person who is familiar with the conversationally-established names, and a second person who is not, the speaker designs the expressions based on the knowledge of the *intended addressee* (Yoon & Brown-Schmidt, 2018). Thus the partner in partner-as-context is the person that the speaker intends to address (not simply whomever happens to be in the same room).

The present research was inspired by a different set of findings that relate to the *retrieval operations* that shape how information is retrieved from memory. In particular, some evidence suggests that context effects are attenuated with more implicit retrieval tasks such as anagram completion or category exemplar generation, compared to more explicit recall tasks (Godden & Baddeley, 1980; Parker, et al., 2007; Mulligan, 2011; Macken, 2002; McKone & French, 2001). For example, in a study by Parker, et al. (2007), participants rated a series of words for their pleasantness in one room, and then were later asked to either *recall* those words (an explicit retrieval task) or to complete a category exemplar task where they were asked to write down words that came to mind in a series of categories (where the categories corresponded to words that had previously been studied – an implicit memory task). The memory task either took place in the same room as study or in a new room, and the two rooms were distinguished by different scents, lighting, music and so forth. Parker, et al. (2007) found a standard contextual reinstatement effect, such that more of the studied words were produced at test when study and test occurred in the same room, compared to a different room. But this context effect emerged only in the explicit recall task. There was no effect of context in the implicit task. We should note, however, as it foreshadows the findings from the present research, that robust effects of context have been reported in some implicit tasks (see Smith, et al., 2018 for a recent example). We will return to this point in the General Discussion. In sum, many studies in the literature find that awareness and explicit recollection enhance context effects in memory, with more consistent context effects found in tasks that emphasize conscious recollection over familiarity based processing of an immediately available stimulus (see Smith, et al., 2018; Parker, et al., 2007; Mulligan, 2011; Godden & Baddeley, 1975; Mackin, 2002). This was the inspiration for the present research.

The present research

The aim of the present research is to test the hypothesis that retrieval processes will affect the partner-specificity of language use. According to the people-as-contexts view of partner specificity in conversation (Brown-Schmidt, et al., 2015), partners are a type of context that shape the retrieval and use of person-relevant information in conversation. If context effects are magnified by explicit retrieval operations, partner effects should be larger when the linguistic information associated with that partner must be more explicitly recalled. Some evidence that is potentially consistent with this hypothesis comes from Yoon & Brown-Schmidt (2018) who report a partner-specific effect in audience design only when there was a delay between the learning of image-label mappings, and the later use of these labels with a familiar vs. a new partner (with the delay potentially resulting in a more difficult retrieval task). In the present work, two experiments test the hypothesis that when labeling an abstract image in a referential communication task, speakers should show a larger partner effect (i.e., greater sensitivity to the partner's [lack of] knowledge) in a case where the to-be-labeled image is not physically present and must be recalled from memory, compared to a case in which the image is clearly visible in the immediate environment. Such a finding would be expected if, when recalling the missing image, speakers use the current addressee as a retrieval cue, thereby increasing attention to the partner in the context, making it more likely that the speaker would take into account the partner's knowledge when describing the image. By contrast, when describing an image in the immediate environment, speakers may be less likely to use the addressee as a retrieval cue as the image can simply be described based on its properties without recalling it from memory.

Experiment 1

Experiment 1 uses a variant of the referential communication task (Krauss & Weinheimer, 1966; 1964) to test the hypothesis that when the label for an image must be explicitly recalled, the more likely the speaker would invoke the current partner as a potential retrieval cue. If so, we would expect to observe greater sensitivity to the partner and thus an amplified audience design effect when retrieval is more explicit.

Method

Participants

The experiment was designed as a 3-person conversation study, with one person assigned to the role of Speaker and the other two assigned to the roles of Matcher A and Matcher B. The focus of the planned analyses is on the performance of the Speaker. Participants were scheduled to participate in the study in groups of 3, though due to recruitment challenges it was anticipated that we would not always have 3 individuals sign up to participate at the same time. When 2 or 3 participants did show up, one person was randomly assigned the role of Speaker and the other person or persons were assigned the role of Matcher. When only one person was available to participate, that person was assigned the role of Speaker. When only 1 or 2 participants were recruited, lab assistants played the Matcher role so that we had a full group of 3 partners. For the two Matchers, assignment of participants and lab assistants to the Matcher A / B roles was random.

A total of 50 native English speaking participants with normal or corrected-to-normal hearing and vision participated as Speakers in this experiment. An additional 92 participants played the role of Matcher A or B (the remaining 8 Matcher A & B roles were played by lab assistants). After the first 10 groups of participants were run we discovered a coding error that irrecoverably scrambled the data files for the Speakers in groups 2-10. This error was fixed and the remaining 40 groups were run without incident (no other changes were made to the script). As a result, we report the data from 41 Speakers.

Materials

The materials were based on Experiment 2 from Yoon and Brown-Schmidt (2018), and consisted of 128 abstract images that were selected to lack a conventional label. A subset of the images were previously used in several other studies (Arnold, Hudson Kam, & Tanenhaus, 2007; Brown-Schmidt, 2009). The images were clustered into sets of four images; in total there were 32 four-image sets across the entire study.

Design

The experimental design was based on previous work using a train-test variant of the classic referential communication task (Yoon & Brown- Schmidt, 2018; 2014). Each group completed 8 rounds of game play where each round was composed of a set of 16 Training trials and a set of 20 Testing trials. As in the prior work, the critical data and focus of our analyses are the Speaker's utterances produced during Testing trials. Each of the 8 rounds featured 16 of the images (four of the 4-image sets), thus across the entire study, each group saw all 128 images.

Training block. During the training block, the Speaker and Matcher A were seated at separate computers in the same room, and worked together to complete a referential communication task. Meanwhile, Matcher B waited outside the room behind a closed door. The Speaker and Matcher A were able to communicate freely using words, however the seating configuration was such that large computer monitors prevented communication using facial and manual gestures. Each training block consisted of a series of 16 sorting trials that were in a set, pre-randomized order. The aim of these trials was for the partners to establish labels for the 16 images in that block. On each trial, the partners saw 4 images from a single image set but in different arrangements on the two screens. The task was for the Speaker to tell Matcher A how to rearrange their images such that Matcher A's were in the same arrangement. Across the 16 sorting trials, the Speaker referenced each of the 16 images four times each. The participants were allowed to speak freely, resulting in the collaborative formation of brief labels for each of the 16 images in that block (Figure 1).

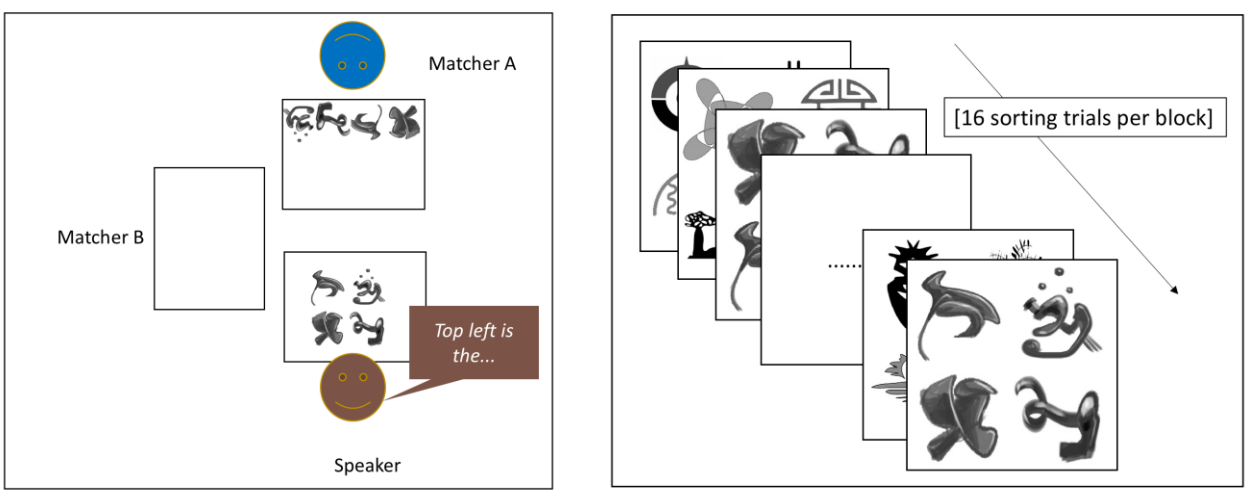

Figure 1. Illustration of the training trials in a single block. Left panel: The Speaker and Matcher A were seated on opposite sides of a table, in front of large computer screens; Matcher B was out of the room. The Speaker and Matcher A viewed the same 4 images on their respective computer screens (but in different arrangements). The task was for the Speaker to instruct Matcher A on how to re-arrange their images by dragging pictures to different parts of the screen such that their screens matched. Right panel: Sequence of training trials.

Testing block. During the testing block, Matcher B was invited back into the room and was seated at a third computer in the room (Figure 1). The testing block consisted of a series of 16 target trials and 4 filler trials. On each trial, Matchers A and B saw one of the 4-image sets that were used in the training block. The images were randomly assigned to one of 4 different locations on the screen. On the Speaker's screen, an instruction at the top of the screen said "Give an instruction to Matcher A" or "Give an instruction to Matcher B". One of the 4 locations had a box around it, and the task was for the Speaker to tell either Matcher A or B to click on the object that was in that location. On the 16 target trials, the target was one of the 16 images in that block. The 4 filler trials were repetitions of four of these 16 target trials and were included in order to prevent participants from using process-of-elimination to figure out what the target was in the later trials (fillers were not intended for analysis). Hence, the target trials were always the first time that the Speaker referenced a given picture at test. The manipulation of retrieval was implemented as follows: On half of the target trials, the target image was present on screen and the Speaker simply had to describe it for the Matcher. On the other half of target trials, the target image was replaced by a "?" mark, and as a result, the Speaker was required to recall the target image from memory, using the 3 other images on screen, along with any other retrieval cues (such as Matcher A from training) to retrieve that image and its label from memory (Figure 2).

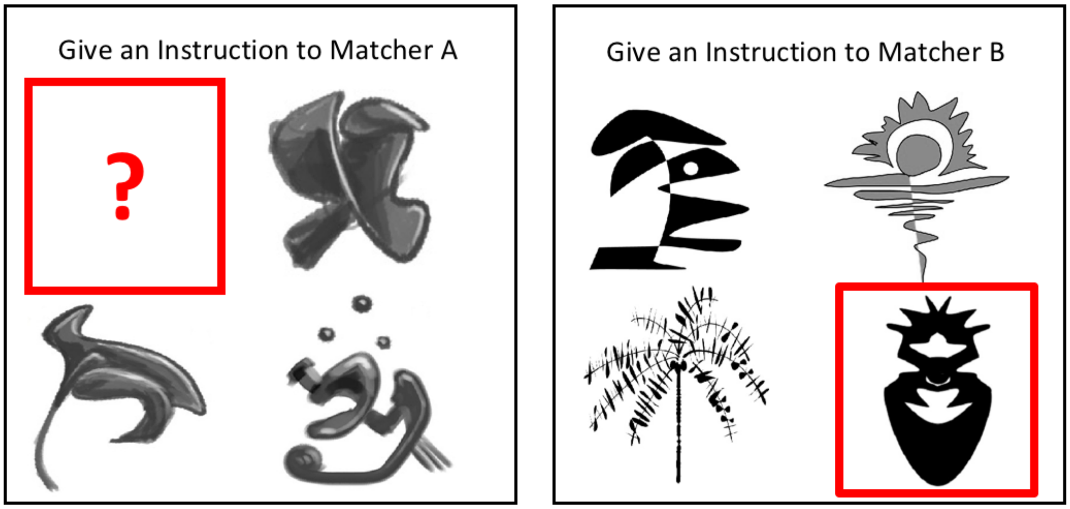

Figure 2. Illustration of the testing trials. The left panel shows a trial where the Speaker addresses Matcher A in the Recall condition. The right panel shows a trial where the Speaker addresses Matcher B in the Describe condition.

Speakers were asked to recall the missing image from memory, and were discouraged from using process of elimination or asking for help from their partner to discover the missing image. In sum, our analyses focus on the speaker's description of the target image on testing trials. The critical manipulations are the addressee at test (Matcher A who participated in the training trials, or Matcher B who did not), and retrieval operation at test (describe vs. recall).

Analysis

The testing trials were transcribed, and the description of the target object on that trial was identified. We coded the number of words used to describe the target image, including all lexical disfluencies (e.g. “um”), function and content words. Lexical disfluencies were included in the word count based on prior work indicating that disfluencies influence online processing (Arnold, Hudson Kam, & Tanenhaus, 2007) and that different forms of disfluency signal different information (Fox Tree, 2001; Fraundorf & Watson, 2014). Words that occurred prior to the beginning of the description that were related to task management (e.g., “the top left one is”) were not included in the word count measure. Lastly, to ensure that this word count measure reflected the Speaker’s audience design process (rather than the listener’s interpretation processes), following prior work, this word count measure only included words produced *prior* to any feedback from the Matcher, also known as the “initiating reference” (see Yoon et al., 2017; Yoon & Brown-Schmidt, 2018; Duff et al., 2011). If the Speaker failed to describe the target before the Matcher intervened, that trial was excluded from analysis of referential form.

For the two targets illustrated in Figure 2, example descriptions from the Speakers, and the description length measure for each description (underlined) are as follows:

Left panel (Recall condition):

Matcher A: “*Matcher A, select the backwards six*” [3 words]

Matcher B: “*Matcher B the animal with uh that has its hind legs but no front legs and the neck is sort of curving in on itself*.” [23 words]

Right panel (Describe condition):

Matcher A: “*For Matcher A, it is the guitar pick at the bottom*” [8 words]

Matcher B: “*Matcher B, this is uh it basically looks like it's a high contrast black and white kind of looks like a cool space alien*” [22 words]

A supplemental analysis of the disfluency rate examined whether the referring expression was fluent or whether it contained one or more disfluencies. A description was coded as disfluent if it contained a lexical disfluency (e.g. “um”, “uh”), lengthened words (e.g., “theee”) or repairs (e.g., “…with the three er, four legs…”).

Lastly, for trials in the Recall condition, we coded whether or not the Speaker successfully recalled the image-label mapping. A trial was coded as having a recall failure when Speakers asserted that they could not recall the image label, as in “*This was the flow-? No I can't remember,*” uttered disfluencies until the Matcher interrupted (e.g. “uh……………….”), or if they worked with their partner through the process of elimination to try and figure out which item was the target, as in “*It's the uh what was it, it's not the Pepsi, not the egg, not thee one that looks like an art piece [Matcher A: the laser.] yeah, the laser. [Matcher A: nice.]*”.

Predictions

A standard audience design effect (Wilkes-Gibbs & Clark, 1992; Yoon & Brown-Schmidt, 2018; 2014) is expected such that Speakers will use many more words to describe the target images when addressing Matcher B, who does not know the image-label mappings, compared to Matcher A. Further, we expect to observe an increased disfluency rate when they were addressing the naïve Matcher B, reflecting the increased utterance design effort in this case (Yoon & Brown-Schmidt, 2018; 2014).

If explicit retrieval operations are more context-sensitive than implicit retrieval operations (Parker et al., 2007), then we would expect to see a stronger influence of the conversational partner in the Recall condition compared to the Describe condition. When addressing Matcher A, in the Recall condition, we hypothesize that Speakers invoke Matcher A as a contextual cue to retrieve the missing image and its label from memory. By contrast, in the Describe condition, Speakers may be less likely to invoke Matcher A as a contextual cue and instead use the image itself as the primary cue to retrieve the image label from memory. If so, the Recall condition should place a greater emphasis on the intended addressee's knowledge as the Speaker designs the referring expression. This predicts that the partner effect (i.e. longer referring expressions when speaking to the naïve Matcher B than the knowledgeable Matcher A) should be larger in the Recall condition compared to the Describe condition (Figure 3). Alternatively, a similarly-sized partner effect across task types would be consistent with the view that partner-referent associations can automatically influence production (Horton, 2007).

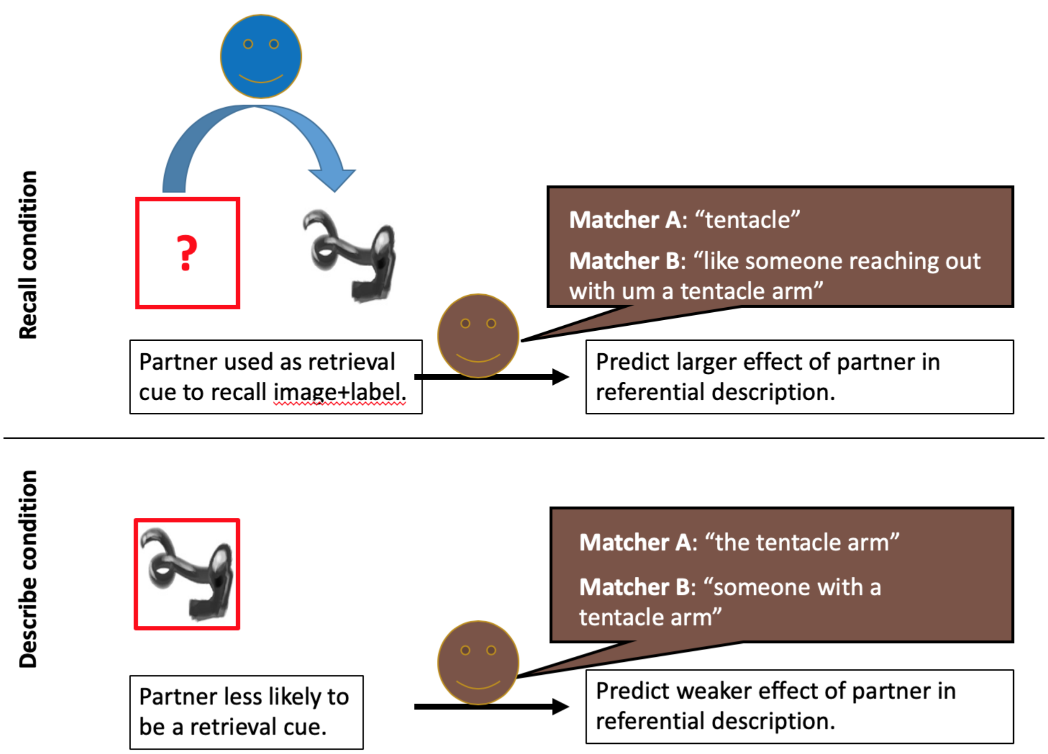

Figure 3. Illustration of the proposed relationship between retrieval processes and audience design at test. In the Recall condition (top) the Speaker is hypothesized to be more likely to consider the conversational partners during the referential process, compared to the Describe condition (bottom), predicting a larger audience design effect in the Recall condition.

Results

This dataset consists of a total of 5248 trials on which the Speaker was tasked with describing a target object to one of the two Matchers (41 Speakers * 8 blocks of trials * 16 target trials per block = 5248). Of those, 20 trials were eliminated from analysis because the Speaker misunderstood the task and did not reference the target, skipped a trial, or addressed the wrong Matcher. Twenty additional trials were excluded because there was no reference to the target prior to the Matcher intervening with feedback. Lastly 7 trials were excluded due to technical issues with the microphone or computer. An additional 1238 trials were excluded from the analyses of referring expression length and disfluency rate because the Speaker was unable to recall the target image label, and as a result there was no referring expression to analyze.

Description length

Following prior work, the analysis of description length is the primary measure of audience design, and we expect Speakers to produce longer descriptions when addressing Matcher B compared to Matcher A (see Yoon & Brown-Schmidt, 2014, 2018). The average description length and disfluency rates are shown in Table 1. Note that there are fewer observations in the Recall condition because the Speaker often failed to recall the image label in that condition.

Experiment 1	**Length**	**SD**	**Disfluency**	***n***
Matcher A - Recall	7.19	5.98	0.34	693
Matcher B - Recall	12.93	9.01	0.52	672
Matcher A - Describe	5.00	3.64	0.13	1303
Matcher B - Describe	9.83	7.05	0.32	1295
Experiment 2				
Matcher A - Recall	9.03	7.57	0.31	680
Matcher B - Recall	11.78	9.10	0.38	719
Matcher A - Describe	8.27	7.39	0.23	1194
Matcher B - Describe	11.23	9.00	0.31	1189

Table 1. Descriptive statistics for Experiments 1 and 2. Average description length (Length = number of words per image; SD=Standard deviation). Disfluency rate by condition (Disfluency).

A log-link mixed effects model was fit to the description length measure using the glmer function in the lme4 (Bates, Mächler, Bolker, & Walker, 2015) package in R (R Core Team, 2016). The fixed effects were coded with orthogonal contrasts. The *Partner* factor compares trials where the Speaker was addressing the knowledgeable Matcher A vs. naïve Matcher B (coded as A = -.5 and B = +.5). The *Retrieval* factor compares trials where the Speaker was describing a visible target vs. when they had to recall and describe that target from memory (coded as Describe = -.5 and Recall as +.5). An initial model with the maximal number of random effects specified by the design failed to converge, and refitting the model with different optimizers using the allFit function (Bates, et al., 2018) revealed inconsistent estimates for the fixed effects. The model was simplified by removing random slopes one-by-one until a final model converged (Barr, et al. 2013). The final model is reported in Table 2.

FIXED EFFECTS	***ESTIMATE***	***SE***	***Z-VALUE***	***P-VALUE***
(INTERCEPT)	1.999	0.046	43.791	<.0001
RETRIEVAL	**0.275**	**0.013**	**21.571**	**<.0001**
PARTNER	**0.645**	**0.056**	**11.429**	**<.0001**
RETRIEVAL*PARTNER	-0.106	0.062	-1.698	0.09
RANDOM EFFECTS	***Variance***	***Std.Dev.***	***Corr***	
ITEM (INTERCEPT)	0.071	0.267		
PARTNER	0.095	0.308	0.23	
RETRIEVAL*PARTNER	0.257	0.507	0.1	0.43
SUBJECT (INTERCEPT)	0.061	0.246		
PARTNER	0.092	0.303	-0.46	
RETRIEVAL*PARTNER	0.045	0.212	-0.11	0.29

Table 2. Experiment 1: Results of a log-link mixed-effects model of description length for the 3963 trials on which the Speaker described the target image, 128 items (images), 41 persons. Bolded values indicate significant effects.

The model of description length revealed main effects of Partner, due to longer expressions with the naïve Matcher B (z = 11.43, p < .0001), and Retrieval type, due to longer expressions in the Recall condition (z = 21.57, p < .0001). The Partner*Retrieval interaction was not significant (z = -1.698, p = .09). Planned comparisons revealed a robust effect of Partner in both the Describe condition (β = .698, z = 12.62, p < .0001), and the Recall condition (β = .592, z = 8.17, p < .0001).

Disfluency Rate

The disfluency rate is analyzed as a supplemental analysis of audience design. The disfluency rate is expected to increase as utterance length increases (Shriberg, 1996), and indeed, disfluent utterances were on average longer (13 words) than fluent utterances (6 words). Thus the data pattern is expected to complement the primary analysis of description length, and the analyses are treated as supplemental to the primary analysis[1]. Following prior work, we expect a higher disfluency rate when the Speaker addresses Matcher B compared to Matcher A, due to the added difficulty in planning and producing a longer description to accommodate B's naiveté (see Yoon & Brown-Schmidt, 2014, 2018). A logit-link mixed effects model was fit to a binary measure of disfluency (0=fluent, 1=disfluent), thus we are modelling the log odds of a disfluent utterance (see Baayen, 2004; Wright, Horry, & Skagerberg, 2009 for discussion). The fixed effects were coded with the same contrasts as before: *Partner* (coded as A = -.5 and B = +.5), and *Retrieval* (coded as Describe = -.5 and Recall = +.5). A model with the maximal number of random effects resulted in convergence errors. A backwards stepping procedure was used to incrementally remove random effects that accounted for the least amount of variance until the model converged (Table 3).

Fixed effects	***ESTIMATE***	***SE***	***Z-VALUE***	***P-VALUE***
(INTERCEPT)	-0.890	0.096	-9.271	<.0001
Retrieval	**1.136**	**0.081**	**13.964**	**<.0001**
Partner	**1.024**	**0.091**	**11.207**	**<.0001**
Retrieval*Partner	**-0.418**	**0.159**	**-2.632**	**<0.01**
Random effects	***Variance***	***Std.Dev.***	***Corr***	
ITEM (INTERCEPT)	0.161	0.401		
PARTNER	0.205	0.453	-0.42	
SUBJECT (INTERCEPT)	0.254	0.504		

Table 3. Experiment 1: Results of a logistic mixed-effects model of the disfluency rate for the 3963 trials on which the Speaker described the target image, 128 items (images), 41 persons.

[1] A post-hoc analysis of the disfluency rate that included description length as a covariate produced a very similar pattern of results, with effects of Partner ($z = 2.49$, $p < .05$), Retrieval ($z = 9.99$, $p < .0001$), and an interaction in the opposite direction as predicted ($z = -2.40$, $p < .05$).

The analysis of disfluencies revealed a main effect of Partner, due to more disfluencies when the Speaker addressed Matcher B ($z = 11.21$, $p <.0001$), and Retrieval type, due to an increase in the disfluency rate in the Recall condition ($z = 13.96$, $p < .0001$). These main effects were qualified by a Partner*Retrieval interaction ($z = -2.63$, $p < .01$). Follow up tests revealed that the effect of Partner was larger in the Describe condition ($\beta = 1.23$, $z = 10.64$, $p < .0001$), compared to the Recall condition ($\beta = .815$, $z = 6.47$, $p < .01$). This interaction is in the *opposite* direction than was predicted.

Memory

As noted above, Speakers frequently failed to recall the name of the missing image in the Recall condition. On average, recall rates were 53.1% when addressing the familiar Matcher A, and 51.7% when addressing Matcher B. Thus in a supplemental analysis, we examined the rate at which Speakers successfully recalled the image-label mappings, and whether this is affected by the identity of the intended addressee.

A logit-link mixed effects model was fit to a binary measure of recall for trials in the Recall condition (failure = 0; recall = 1). The model included *Partner* as a fixed effect (coded as A = -.5 and B = +.5). A model with the maximal number of random effects specified by the design failed to converge. A backwards stepping procedure was used to incrementally remove random effects that accounted for the least amount of variance until a model that did not result in warnings was identified (Table 4).

FIXED EFFECTS	ESTIMATE	SE	Z-VALUE	P-VALUE
(INTERCEPT)	0.183	0.194	0.943	0.346
PARTNER	-0.080	0.098	-0.816	0.415
RANDOM EFFECTS	***Variance***	***Std.Dev.***	***Corr***	
ITEM (INTERCEPT)	1.392	1.180		
PARTNER	0.096	0.310	-0.350	
SUBJECT (INTERCEPT)	0.998	0.999		

Table 4. Experiment 1: Results of logit-link mixed effects model on retrieval of the image label. 2603 observations, 41 participants, 128 items.

Results of the model indicated that the partner whom the Speaker was addressing did not significantly affect whether or not Speakers would recall the image labels (z = -.82, p = .42).

Discussion

The findings point to a clear audience design effect: when the Speaker was addressing Matcher A, who shared knowledge of the conversationally-established image-label mappings, the Speaker used brief, conversationally-established labels. When addressing Matcher B, who was not familiar with the image-label mappings, the Speaker produced significantly longer expressions that were more likely to be disfluent. Both Matchers A and B were present in the room at test and the Speaker randomly alternated between addressing Matcher A vs. Matcher B at test. As a result, this audience design effect arises from the Speaker's intention to address either Matcher A or Matcher B on that particular test trial. This finding, that referential form is shaped by the intended addressee in multiparty conversation, replicates recent findings (Yoon & Brown-Schmidt, 2018; 2019).

The results of Experiment 1 were inconsistent with the hypothesis that audience design effects are magnified with more explicit (vs. implicit) retrieval operations. The interaction between

Partner and Retrieval operation was significant only in the analysis of disfluency rate, and was in the opposite direction as predicted. This result suggests that if anything, audience design effects were larger when retrieval of the image-label mapping simply involved describing a visible picture. This finding, while a surprise, is potentially consistent with arguments that retrieving partner-specific information from memory is difficult, and attenuated when speakers are under time pressure (Horton & Keysar, 1996) or when working memory resources are limited (Wardlow, 2013).

Of note is that speakers did tend to produce longer expressions that were more likely to be disfluent in the recall condition. These effects likely reflect the more difficult task of having to bring to mind the missing target image in this condition. Additionally, speakers in this condition may have experienced a feeling that they knew the missing image (Smith & Clark, 1993; Tullis & Fraundorf, 2017), but could not bring it to mind, reflected in lengthy expressions, e.g. " thee.. ok so I know which one this is….".

Experiment 2

The aim of Experiment 2 was to replicate the findings of Experiment 1. In order to provide a more uniform experience for participants in Experiment 2, we recruited participants to play the role of Speaker only. Lab assistants played the roles of Matchers A and B (across participants the lab assistants took turns playing the role of Matcher A and B so that a particular lab assistant wasn't always in one of the two roles).

Method

Participants

The experiment was designed as a 3-person conversation study. Participants were assigned the role of Speaker and completed the study with two lab assistants who played the roles of Matcher A and Matcher B. As in Experiment 1, the focus of the planned analyses is on the performance of the Speaker. A total of 40 native English speaking participants with normal or corrected-to-normal hearing and vision participated as Speakers in this experiment. An additional 5 participants were run in the study but their data was not analyzed because equipment failure caused their data to not be saved.

Materials, Design, Analysis

The materials were identical to those in Experiment 1 and consisted of 128 abstract images that were selected to lack a conventional label. The experimental design was identical to Experiment 1 with the following exceptions: While the training block was identical to Experiment 1, the testing block was slightly different. The structure of test trials was the same as in Experiment 1, except that in order to discourage participants from using a process of elimination to identify the target (e.g. *Not the bear, not the falling leaves…*) we added a "forget" button at the bottom of each person's screen that would allow them to move forward to the next test trial if the Speaker could not remember what the target was. This change was made because we discovered that in Experiment 1, when Speakers could not remember the name of the target image, they sometimes engaged in a process of elimination that took a long time to complete. The analysis of Experiment 2 was identical to that in Experiment 1. Occasionally, the participant clicked the "forget" button when they did in fact successfully recall the target image (occurring

<1% of the time). Use of the “forget” button was not considered in the analysis of memory data as we elected to use the identical coding system as in Experiment 1, where a response was treated as a recall only if the participant successfully recalled the target image label without assistance from the conversational partner.

Predictions

As in Experiment 1, we anticipated a standard audience design effect (Wilkes-Gibbs & Clark, 1992; Yoon & Brown-Schmidt, 2018; 2014) with Speakers using more words (and more often being disfluent) when addressing the naïve Matcher B. Given the results of Experiment 1 (which were *not* as predicted), we did not expect to find increased audience design effects in the Recall condition.

Results

By design the dataset would consist of a total of 5120 trials on which the Speaker was tasked with describing a target object to one of the two Matchers (40 Speakers * 8 blocks of trials * 16 target trials per block = 5120). A portion of the data from 7 participants was lost due to computer error (n=1), or because the participant did not finish the study within the two hour period (n=6), removing 304 trials from the analysis. 32 further trials were removed from the analysis due to procedural errors. The remaining 4784 trials were submitted to analysis, including 3782 trials on which the Speaker successfully described a target image.

Description length

The analysis of description length is the primary measure of audience design, and we expect Speakers to produce longer descriptions when addressing Matcher B compared to Matcher A.

The average description length and disfluency rates for Experiment 2 are shown in Table 1. There are fewer observations in the Recall condition because the Speaker often failed to recall the missing target image in that condition.

A log-link mixed effects model was fit to the description length measure as before. The *Partner* factor compares trials where the Speaker was addressing the knowledgeable Matcher A vs. naïve Matcher B (coded as A = -.5 and B = +.5). The *Retrieval* factor compares trials where the Speaker was describing the target vs. when they had to recall that target from memory (coded as Describe = -.5 and Recall as +.5). The model was fit with the maximal number of random effects specified by the design (Table 6).

Fixed effects	**Estimate**	**SE**	**z-value**	**p-value**	
(Intercept)	2.176	0.073	29.637	<.0001	
Partner	**0.304**	**0.050**	**6.066**	**<.0001**	
Retrieval	0.057	0.033	1.721	0.09	
Partner*Retrieval	-0.033	0.054	-0.611	0.54	
Random effects	***Variance***	***Std.Dev.***	***Corr***		
Item (intercept)	0.078	0.279			
Partner	0.043	0.208	0.050		
Retrieval	0.040	0.201	0.050	-0.040	
Partner*Retrieval	0.160	0.400	-0.150	0.310	0.260
Participant (intercept)	0.189	0.435			
Partner	0.080	0.282	-0.210		
Retrieval	0.024	0.155	-0.360	0.120	
Partner*Retrieval	0.038	0.195	0.040	-0.130	0.100

Table 6. Experiment 2: Results of a mixed-effects model of description length for the 3782 trials on which the Speaker described the target image, 128 items (images), 40 persons.

The model of description length revealed a main effect of Partner, due to longer expressions with the naïve Matcher B (z = 6.07, p < .0001). While descriptions were somewhat longer in the Recall condition, the effect of Retrieval was not significant (z = 1.72, p =.09). The interaction was not significant (p =.54).

Disfluency Rate

The disfluency rate is analyzed as a supplemental analysis of audience design, as the disfluency rate is expected to pattern with description length. As in Experiment 1, disfluent expressions were on average longer (15 words) than fluent expressions (8 words). We expected a higher disfluency rate when the Speaker was addressing Matcher B compared to Matcher A. A logit-link mixed effects model was fit to a binary measure of disfluency. The fixed effects were coded with the same contrasts as before. A model with the maximal number of random effects specified by the design failed to converge. A backwards stepping procedure was used to incrementally remove random effects that accounted for the least amount of variance until a model that did not result in convergence warnings was identified (Table 7).

Fixed effects	Estimate	SE	z-value	p-value
(Intercept)	-1.016	0.165	-6.172	<.0001
Partner	**0.465**	**0.118**	**3.952**	**<.0001**
Retrieval	**0.408**	**0.083**	**4.929**	**<.0001**
Partner*Retrieval	-0.091	0.180	-0.505	0.61
Random effects	Name	Variance	Std.Dev.	
Item (intercept)	0.261	0.511		
Partner*Retrieval	0.536	0.732	0.120	
Participant (intercept)	0.918	0.958		
Partner	0.258	0.508	-0.060	

Table 7. Experiment 2: Results of a logistic mixed-effects model of the disfluency rate for the 3782 trials on which the Speaker described the target image, 128 items (images), 40 persons.

The analysis of disfluencies revealed a main effect of Partner, due to more disfluencies produced when the Speaker addressed Matcher B ($z = 3.95$, $p < .0001$), and Retrieval type, due to an increase in the disfluency rate in the Recall condition ($z = 4.93$, $p < .0001$). The interaction was not significant ($p = .61$)[2].

Memory

In Experiment 1, Speakers were equally likely to recall the image label when addressing Matchers A and B. For Experiment 2, a logit-link mixed effects model was fit to a binary measure of recall (recall = 1; failure = 0). The fixed effects were coded with the same contrasts as before. A model with the maximal number of random effects specified by the design failed to converge. A backwards stepping procedure was used to incrementally remove random effects

[2] A post-hoc analysis of the disfluency rate that included description length as a covariate again revealed an effect of Retrieval (z = 4.06, p < .0001), and a non-significant interaction (z = 0.06, p = .95). Unlike the planned analysis, the effect of Partner was no longer significant (z = -0.38, p = .70), which is unsurprising as description length and disfluency rates clearly pattern together in the data.

that accounted for the least amount of variance until a model that converged was identified (Table 8).

FIXED EFFECTS	ESTIMATE	SE	Z-VALUE	P-VALUE
(INTERCEPT)	0.448	0.159	2.815	<.01
PARTNER	0.165	0.096	1.717	0.086
RANDOM EFFECTS	*Variance*	*Std.Dev.*	*Corr*	
ITEM (INTERCEPT)	0.994	0.997		
PARTICIPANT (INTERCEPT)	0.605	0.778		
PARTNER	0.012	0.111	-1.000	

Table 8. Experiment 2: Results of logit-link mixed effects model on image label recall for the 2401 trials in the Recall condition, 128 items, 40 persons.

Speakers successfully recalled the image labels 57% of the time when addressing the familiar Matcher A, and 60% of the time when addressing Matcher B. This difference, which is in the opposite direction as predicted, was not significant ($p = .09$).

General Discussion

Two experiments were conducted to test a prediction derived from the people-as-contexts view of conversation (Brown-Schmidt, et al., 2015). Inspired by findings in the memory literature that the effect of reinstating an environmental context is enhanced when memory retrieval is more explicit (Godden & Baddeley, 1980; Parker, et al., 2007; Mulligan, 2011), it was hypothesized that sensitivity to the partner-as-context would be heightened when the target image had to be explicitly recalled from memory, rather than described based on information in the immediate context.

The results of two experiments did not support this hypothesis. In both experiments, we observed a robust audience design effect, such that speakers used more words, and were more frequently disfluent when addressing the new partner who lacked knowledge of the previously established referential labels for the images. This audience design effect replicates prior findings that referential form is influenced by whom the speaker intends to address in multiparty settings (Yoon & Brown-Schmidt, 2018; 2019). This shows that speakers distinguished between the two addressees, and successfully designed a referring expression that was appropriate for the intended addressee. However, there was no evidence that audience design was magnified in the more explicit recall task. If anything, the partner effect was *smaller* when speakers had to explicitly recall the target image.

We also observed that speakers were *not* more likely to correctly recall the target image when addressing the more knowledgeable partner. This lack of an effect of the partner on image label recall is reminiscent of the finding that partners did not significantly affect the speed of lexical retrieval in a picture naming task (Brown-Schmidt & Horton, 2014; cf. Horton, 2007). One interpretation of this finding is that while partners serve as a strong cue to designing an appropriate referential description (i.e. an audience design effect), partners only weakly shape retrieval of the relevant to-be-communicated information (such as the characteristics of a referent, including previously-established labels for that referent) in the first place.

The disconnect between the present results and prior findings in the contextual reinstatement literature may owe to the different tasks involved. For example, the contextual reinstatement effects examined in Parker, et al. (2007) concerned effects of the same vs. different room on word recall vs. category exemplar generation. It may be that these processes are too distinct from the recall and description tasks used in the present research for analogous effects to hold. It may also be the case that paradigms in which retrieval of partner-associated information is more difficult or the partners less distinct (Horton & Gerrig, 2005; Horton & Slaten, 2012) would be more likely to reveal effects of retrieval processes. While speakers frequently failed to recall the missing target image, information about which partner was familiar with that missing image was trivial to access (it was always Matcher A), thus a simple "one-bit" model of which partner is knowledgeable in the setting (Galati & Brennan, 2010) could have supported audience design. Further, in both experiments, speakers had repeatedly produced the image labels, a generation process that is known to boost the accessibility of that information in memory (Knutsen & LeBigot, 2014; 2017; McKinley, et al., 2017; MacLeod, et al 2010). Perhaps, then, in a design in which different partners were associated with different images (e.g. Horton & Gerrig, 2005), making it more difficult to recall which partner knows which image labels, retrieval processes would play a stronger role in audience design. Retrieval processes may also play a larger role in audience design if the participant were focused more directly on the intended addressee rather than the to-be-described pictures.

A final consideration is that the lack of an enhanced partner-as-context effect in the more explicit retrieval task is potentially consistent with findings that context can guide automatic retrieval of information from memory in some tasks (see Smith, et al., 2018; Vakil, et al. 2007), as well as evidence from Smith and Vela's (2001) meta-analysis which reported a physical reinstatement effect size that was similar across several types of explicit tests (Recall: d=.29; Recognition=.27; Cued Recall=.25). It may be that one's conversational partner provides a strong enough cue to retrieve associated information from memory that retrieval is effective across a range of conversational actions that vary in the automaticity of the associated retrieval processes. While the differences in tasks between, say, anagram completion and picture description in unscripted conversation, prevent point-by-point comparisons, the parallel between the present findings and evidence of pervasive contextual reinstatement effects in memory paradigms (e.g., Smith, et al., 2018; Vakil, et al. 2007; Smith & Vela, 2001) is striking. Moreover, the fact that we observe context (partner) sensitivity in an ecologically valid unscripted conversational task, points to the generality of this type of context effect – here, where one's partner is the context that guides what speakers will say.

In this light, we note an intriguing link between the present findings and prior work examining referential processes in patients with hippocampal amnesia. In a surprising finding, Duff, et al. (2006; 2011) tested patients with bilateral hippocampal damage in a referential communication task and report that they successfully developed and retained novel labels for abstract images despite severe episodic memory impairment. Inspired by these findings, Yoon, Duff, and Brown-Schmidt (2017) asked whether this preserved learning reflected partner-specific common ground. Yoon, et al. (2017) repeated the study in a two-partner paradigm and found that patients with amnesia produced longer expressions when speaking to a partner who lacked knowledge of the image-label mappings. This partner-specific use of language emerged despite an inability to explicitly recall the events that resulted in those image labels becoming shared knowledge. The lack of an effect of retrieval processes in the present findings is largely consistent with these findings: Regardless of how partner-specific information is retrieved from memory, language use is apparently highly tailored to one's conversational partner. This finding is consistent with the idea that multiple memory systems contribute to the use of common ground in healthy persons (Brown-Schmidt & Duff, 2016; Duff & Brown-Schmidt, 2017). The persistence of the partner-specific effects observed in the present research, and in these prior studies may relate to the fact that in each case, the communicative partner was a relevant participant in the communicative exchange. By contrast, in tasks where the partner is less relevant (see Ostrand & Ferreira, 2019; Yoon, et al. 2012), or partner-specific information is in competition with much stronger probabilistic cues (Ryskin, Kurumada, & Brown-Schmidt, 2019), partner-specific effects may be attenuated.

In conclusion, the present findings show that we encode information about past experiences with conversational partners and use this information to design linguistic utterances in a way that is sensitive to the shared past. This allows speakers to describe hard to name images using brief conversationally-developed labels when addressing the partner who shared the experience of developing those labels, and to produce longer descriptions for partners who are unfamiliar with the labels. This partner sensitivity was similar in magnitude across tasks that varied in how the image and its label were retrieved from memory. The fact that audience design effects were persistent across different types of retrieval processes speaks to the ubiquity with which language use in conversation is tailored to the particular people with whom we converse.

Open Practices Statement

The materials for all experiments are available from SBS upon request. The raw data are available at https://osf.io/qsk4u/.

Acknowledgements

This material is based on work supported by National Science Foundation Grants NSF BCS 1556700 and 1921492 to Sarah Brown-Schmidt. Thank you to Casey O'Quinn and many other dedicated research assistants for help running the studies and coding the data.

References

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

Baayen, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental lexicon working papers*, *1*(1), 1-47.

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*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 482–493.

Brown-Schmidt, S., Yoon, S. O., & Ryskin, R. A. (2015). People as Contexts in Conversation. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 59–99). Academic Press: Elsevier Inc.

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. & Duff, M. C. (2016). Memory and common ground processes in language use. *Topics in Cognitive Science*, *8,* 722-736.

Clark, H. H. (1992). Arenas of language use. University of Chicago Press. Clark, H. H. (1996). Using language. Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In *Advances in psychology* (Vol. 9, pp. 287-299). North-Holland.

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

Craycraft, N.N. & Brown-Schmidt, S. (2018). Compensating for an inattentive audience. *Cognitive Science, 42,* 1504-1528.

Duff, M. C., Gupta, R., Hengst, J. A., Tranel, D., & Cohen, N. J. (2011). The use of definite references signals declarative memory: Evidence from patients with hippocampal amnesia. *Psychological Science*, *22*(5), 666-673.

Duff, M.C., & Brown-Schmidt, S. (2017). Hippocampal contributions to language use and processing. In Hannula, D. & Duff, M.C. (Eds.). *The hippocampus from cells to systems: Structure, connectivity, and functional contributions to memory and flexible cognition* (pp. 503-536). Springer, Cham.

Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 764–770.

Ferreira, V. S. (2019). A mechanistic framework for explaining audience design in language production. *Annual review of psychology*, *70*, 29-51.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, *29*(2), 320-326.

Fraundorf, S. H., & Watson, D. G. (2014). Alice's adventures in um-derland: Psycholinguistic sources of variation in disfluency production. *Language, Cognition and Neuroscience*, *29*(9), 1083-1096.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee?. *Journal of Memory and Language*, *62*(1), 35-51.

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

Godden, D., & Baddeley, A. (1980). When does context influence recognition memory? *British Journal of Psychology, 71*, 99–104.

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*(3), 843.

Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 91.

Horton, W. S., & Brennan, S. E. (2016). The role of metarepresentation in the production and resolution of referring expressions. *Frontiers in psychology*, *7*, 1111.

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

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

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.

Horton, W. S., & Slaten, D. G. (2012). Anticipating who will say what: the influence of speaker-specific memory associations on reference resolution. *Memory & Cognition, 40*, 113–126.

Horton, W. S., & Spieler, D. H. (2007). Age-related differences in communication and audience design. *Psychology and Aging, 22*, 281–290.

Horton, W. S., & Gerrig, R. J. (2016). Revisiting the Memory-Based Processing Approach to Common Ground. *Topics in cognitive science*, *8*(4), 780-795.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language*, *32*(4), 421-445.

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

Knutsen, D., & Le Bigot, L. (2017). Conceptual match as a determinant of reference reuse in dialogue. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 350.

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.

Krauss, R. M., & Weinheimer, S. (1964). Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, *1*(1-12), 113-114.

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 153.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 671–685. http://dx.doi.org/10.1037/a0018785.

McKone, E., & French, B. (2001). In what sense is implicit memory “episodic”? The effect of reinstating environmental context. Psychonomic Bulletin & Review, 8, 806–811. http://dx.doi.org/10.3758/BF03 196221

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*, 201–213.

Mulligan, N. W. (2011). Conceptual implicit memory and environmental context. *Consciousness and cognition, 20*, 737–744.

Ostrand, R., & Ferreira, V. S. (2019). Repeat after us: Syntactic alignment is not partner-specific. *Journal of Memory and Language*, *108*, 104037.

Parker, A., Dagnall, N., & Coyle, A. M. (2007). Environmental context effects in conceptual explicit and implicit memory. *Memory, 15*, 423–434.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved November 11, 2017, from https://www.R-project.org/.

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

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55). Academic Press.

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

Shriberg, E. (1996, October). Disfluencies in switchboard. In *Proceedings of International Conference on Spoken Language Processing* (Vol. 96, pp. 11-14). Philadelphia, PA: IEEE.

Smith, S. M. (1979). Remembering in and out of context. Journal of Experimental Psychology: Human Learning and Memory, 5, 460–471. http://dx.doi.org/10.1037/0278-7393.5.5.460

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of memory and language*, *32*(1), 25-38.

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

Smith, S. M., Handy, J. D., Hernandez, A., & Jacoby, L. L. (2018). Context specificity of automatic influences of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(10), 1501-1513. http://dx.doi.org/10.1037/xlm0000523.

Stalnaker, R. C. (1978). Assertion. *Syntax and Semantics, 9*, 315–332.

Tullis, J. G., & Fraundorf, S. H. (2017). Predicting others' memory performance: The accuracy and bases of social metacognition. *Journal of Memory and Language*, *95*, 124-137.

Vakil, E., Raz, T., & Levy, D. A. (2007). Multifactorial context effects on visual recognition memory. *The Quarterly Journal of Experimental Psychology*, *60*(7), 916-923.

Wardlow-Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speaker's control over leaking private information during language production. *Psychological Science, 17*, 273–277.

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic bulletin & review*, *20*(4), 766-772.

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

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior research methods*, *41*(2), 257-267.

Yoon, S.O. & Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cognitive Science, 43,* DOI: 10.1111/cogs.12774*.*

Yoon, S.O., & Brown-Schmidt, S. (2018). Aim Low: Mechanisms of audience design in multiparty conversation. *Discourse Processes, 55(7), 566-592.*

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

Yoon, S.O., Duff, M.C., & Brown-Schmidt, S. (2017). Learning and using knowledge about what other people do and don't know despite amnesia. *Cortex, 94,* 164-175.