Memory Representations Supporting Speakers' Choice of Referring Expression: Effects of Category Overlap and Shared Experience

Kristen Skovbroten Gorman (ksgorman@bcs.rochester.edu)
Whitney Gegg-Harrison (wgharrison@bcs.rochester.edu)
Chelsea R. Marsh (chelsea.marsh@rochester.edu)
Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Rochester, NY 14607 USA

Abstract

Speakers must take their addressee's knowledge into account in choosing to refer to an object using a name or a description. Do speakers keep track of partner-specific information about the common ground status of names? And if so, what mechanisms support this ability? We present a series of experiments that investigate the nature of the memory representations involved in supporting speakers' ability to distinguish shared from privileged information. The results of these experiments suggest that category information can be used as a cue to aid retrieval of ground status, and that shared experience plays an important role in helping speakers to distinguish privileged information from shared information.

Keywords: language production; common ground; memory representations; referring expressions; shared experience.

Introduction

One of the most basic things we do with language is use it to refer to a particular object or individual in the world. In order to successfully refer, speakers must choose a referring expression that is likely to be understood by their addressees. One of the decisions that speakers face is whether to use a name (e.g. *Inta*) or a description (e.g. *the yellow thing that looks like a worm*). A name will be the shorter and more precise referring expression, but only if the addressee knows the name.

The distinction between names and descriptions has been especially significant within the domains of philosophy of language, semantics, and pragmatics, due to debates about the nature of their meanings and the means by which they refer. Though some (e.g. Searle, 1950) have argued that names are simply shorthand for sets of descriptions and thus refer similarly, more widely accepted accounts (following Kripke, 1980) hold that names like *John* do not carry any descriptive content which addressees could use to ascertain the reference; instead, they refer directly, by virtue of having a special connection to the referent. Thus a name tells an addressee *which* thing is being picked out by the speaker, but does not tell them anything *about* it.

In their account of reference in interactive dialog, Clark and Wilkes-Gibbs (1992) describe the process involved in successful reference as a collaboration between speaker and addressee. They examine conversations in which speakers and addressees must work together to arrange a set of cards printed with complex "tangram" figures, and find that over the course of these conversations, speakers and addressees seem to work together to establish mutually acceptable referring expressions, often starting with descriptions and eventually entraining upon shorter, more name-like expressions. Subsequent studies have shown that these shortened expressions are difficult to understand for individuals who did not participate in the collaborative process, lending further credence to the view that successful reference requires coordination during conversation (Clark & Shober, 1992).

Notions of common ground, coordination, and perspective-taking are central to understanding reference in interactive conversation, and for developing dialog systems that can successfully interact with human users using natural language (e.g. Brennan 2000), and are particularly important for understanding the difference between a name and description. In order to understand a *name*, the addressee needs to know that the name is linked to the particular referent, and in order to felicitously use a name to refer to an object, the speaker must believe that their addressee has that knowledge; thus, in order to successfully refer using a name, interlocutors must have some way of establishing and representing what knowledge is shared with their conversational partner. Clark and Marshall (1978) provide an account of this process in which interlocutors build up detailed representations of shared knowledge over the course of conversation; speakers then use these detailed "reference diaries" in deciding what to say. However, many have argued that the process of building up and maintaining such detailed representations may be too computationally intensive to realistically account for what happens during ordinary conversation (e.g. Keysar, Lin & Barr, 2003).

Because speakers' use of names provides a window into their beliefs regarding their addressee's knowledge, names are a valuable tool for investigating our abilities for perspective-taking and representing common ground. Recent work involving name usage has begun to explore the limits on these abilities, and what those limits might tell us about the nature of the processes and representations involved. For example, Wu and Keysar (2007) argue that instead of tracking whether or not individual pieces of knowledge are shared with a particular addressee, speakers might instead use a global Information Overlap Heuristic. Using such a heuristic, speakers could safely assume that they could rely on their own knowledge when the overlap in

information between themselves and their addressee is extensive; the overall amount of information shared by a speaker and addressee can serve as a cue to whether a particular item is shared. In their experiment, pairs of naïve participants learned novel names for novel shapes. They either had high information overlap and learned 18 of the 24 names together or low overlap and learned only 6 of the 24 names together. The remaining names were taught only to the speaker, making them privileged information. During a subsequent matching task in which speakers had to instruct their partner to select a target shape from an array of three shapes, speakers used substantially more names (rather than descriptions) when a name was shared than when it was privileged. Speakers also used more names overall in the high overlap condition, which Wu and Keysar interpreted as support for their Information Overlap Heuristic.

Heller, Gorman and Tanenhaus (in press) used the same paradigm and replicated the basic Wu and Keysar findings, but a more detailed analysis revealed a difference in the form of the utterances speakers used when names were shared versus when they were privileged. Heller et al. found that when speakers used names for shared shapes, they primarily used a Name-Alone form (a name without any description) and on the small portion of trials where speakers used names for privileged shapes, they primarily used a Name-then-Description form (where a speaker uses a name immediately followed by a description). Their analyses suggest that this Name-then-Description form is not a repair but rather the utterance is planned with the intention to include the description. Therefore, this difference in form suggests that speakers *can* track whether or not an individual piece of information is shared or privileged with respect to their conversational partner, and can use this partner-specific information during production.

Likewise, evidence from comprehension studies suggests that addressees can use speaker-specific information when comprehending words or referring expressions (e.g. Creel, Aslin & Tanenhaus, 2008; Metzing & Brennan, 2003; Brennan & Hanna, 2009). Creel et al. (2008) showed sensitivity to speaker-specific lexical representations in an experiment where participants were exposed to pairs of cohort competitors (e.g. *candle* and *candy*). Members of each cohort pair were spoken either by the same speaker or by different speakers. In the second half of the experiment, participants showed less competition between the target and its cohort competitor when each member of the cohort pair was spoken by a different speaker.

Brennan and colleagues (Metzing & Brennan, 2003; Brennan & Hanna, 2009) have shown that addressees also use speaker-specific information when comprehending referring expressions. In their experiments, participants played a referential communication game where the speaker (a confederate) used consistent referring expressions for particular items. Later, participants played the referential communication game again with either the same speaker or a different speaker. Participants were slower to identify the target and searched the display more when the original partner used a referring expression that was different than the one used in the first part of the study, suggesting that participants store partner-specific information about the use of particular referring expressions.

But if detailed representations of common ground are too computationally expensive to maintain, then how could interlocutors store and use partner-specific information to the extent these studies suggest? Galati and Brennan (2010) suggest that simple "one-bit" representations of partner knowledge may underlie these abilities; if information about the addressee's knowledge is readily available, it can be easily used. Horton and Gerrig (2005) have proposed that information about common ground may be represented as a by-product of ordinary memory processes, which contain context-specific episodic traces, rather than being separately tracked and represented. They suggest that when speakers want to refer to something, they activate episodic memories for that object. If the episodic memory links the name for the referent with their addressee, the speaker uses that name. Using a paradigm in which participants are given training that leads them to associate particular item categories with particular partners, they find stronger effects of audience design when partner-specific associations were easier to distinguish in memory (Horton & Gerrig, 2005b). Horton (2007) provides further evidence that these sorts of effects do not depend on *explicit* recall of a partner-specific link.

In the current series of experiments, using a paradigm similar to Wu and Keysar (2007) and Heller et al. (in press), we investigate the nature of the memory representations that might underlie speakers' ability to distinguish shared from privileged information. In Experiment 1, we manipulate the categorical structure of the information and find modest effects of category structure on speakers' choice of referring expression. In Experiment 2, we examine the role of partner-specific shared experience, and find that speakers have more difficulty using felicitous referring expressions without the shared experience of learning the material with their addressee.

Experiment 1

Experiment 1 examined whether speakers use categorical knowledge when evaluating the ground status of a referent. If speakers use categorical knowledge to aid retrieval, the Horton and Gerrig model predicts that names may be used infelicitously when some knowledge in a category is shared while some is privileged. Additionally, when an entire category is shared or privileged, this category information could improve speakers' ability to distinguish shared and privileged names, since the category of a particular item could serve as a reliable cue to its ground status.

Method

Participants Thirty pairs of native English speakers from the University of Rochester community were paid for their participation. Each pair knew each other prior to the experiment and chose to participate together. One participant was randomly assigned the role of "director",

and the other the role of "matcher". Three additional pairs were excluded from analysis due to equipment failure.

Materials Eighteen novel clipart images of monsters and 6 of robots from clipart.com were used for the training phase of this experiment (see Figure 1 for sample items). Each was randomly assigned an artificial name (e.g. *Grampent, Molget*) from those used by Wu and Keysar (2007). An additional 30 images of monsters and robots were used as distractor images or for practice trials during testing.

The distribution of monsters and robots across shared and privileged ground varied between conditions. In the mixed-category condition, monsters and robots were evenly distributed between shared and privileged ground with 9 monsters and 3 robots in shared ground and 9 monsters and 3 robots in privileged ground. In the shared-category condition, all 6 robots were shared, and most monsters (12 of 18) were privileged. In the privileged-category condition, all robots were privileged and most monsters were shared. When robots and monsters were both in privileged or shared ground, the items were interleaved within training blocks.

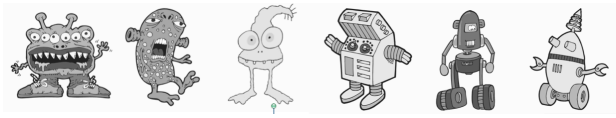

Figure 1. Sample of monsters (left) & robots (right).

Procedure

Training phase The names of the 24 items were taught using 5"x8" flashcards of each item in four blocks of six items each. During the first two blocks, the two participants sat together across from the experimenter and during the third and fourth blocks, the director learned the remaining names with the experimenter while the matcher played a non-linguistic computer game while listening to instrumental music over headphones.

For each item, the experimenter presented a card showing the picture of that item, said the name, and waited for both participants to repeat the name before proceeding to the next card. After going through the six items in the block twice, the experimenter just presented the card and waited for the participants to name the shape. The experimenter then said the name or corrected any errors if the participants could not name the item correctly. The experimenter repeated this procedure until both participants could name all six items flawlessly, and then moved on to the next block.

After all of the blocks had been learned, the experimenter had the participants name the shapes in each block a final time before proceeding to the testing phase.

Testing phase During the testing phase, participants played a referential communication game. Participants sat in front of two different computers and were free to converse over a network, but they could not see each other. The director was presented with one item: Shared Monster, Privileged Monster, Robot (either shared or privileged) or New (half monsters and half robots). The director instructed the matcher who saw three items, to click the target item "as quickly and accurately as possible". The matcher's display contained the target item, another named item from the training set and one unnamed item from the set of distractor items.

Trials were advanced when the matcher clicked on any item: if the matcher clicked the wrong item, an error sound was heard, but participants could not correct the error. The referential communication task had two practice trials followed by 24 experimental trials, six of each type (Shared Monster, Privileged Monster, Robot, New). The same test items were presented for all training conditions.

The participants' utterances were recorded to a computer and later transcribed by the experimenter.

Post-tests After completing the referential communication task, the speaker completed two additional tasks. First, the director was presented with each of the 24 items she had learned during training and had to determine whether both she and her partner learned the name of that item or if only she learned it. The computer presented the items one at a time in random order, and the director had to click "learned together" or "learned alone".

The director then completed a task where she had to name each of the shapes. The computer presented the items one at a time in random order, and the director had to say its name, which was recorded. A response was coded as correct when no more than one phoneme was incorrect.

Results

Task performance Overall **t**ask performance was excellent. Matchers clicked on the correct item on 99% of the trials. Directors' accuracy on the first post-test, which tested their ability to explicitly distinguish between items they and their partners both learned and which they learned alone, was 97% with no significant differences between the training conditions.[1] Director's accuracy on the second post-test, which tested their knowledge for the names of the items, was 76% (70% for privileged-category, 75% for shared-category, 84% for mixed) with no significant differences between the training conditions.

Directors' Utterances We assigned the directors' first turn of each trial to one of six categories: No-Name (a description without using a learned name), Name-in-Description (a shared name used non-referentially as part of the description; e.g. *looks like Grampent*), Description-then-Name (a description followed by a name), Name-then-Description (a name followed by a description), Knowledge-Query (asking if their partner knows the name) and Name-Alone (a name without any description of the item). Our analyses will focus on the Name-Alone form since Heller et al. (in press) showed that speakers use

[1] To test significance for task performance, we conducted one-way ANOVAs on quasi-logit transformed proportions and for Experiment 2 we conducted two-tailed paired t-tests on quasi-logit transformed proportions.

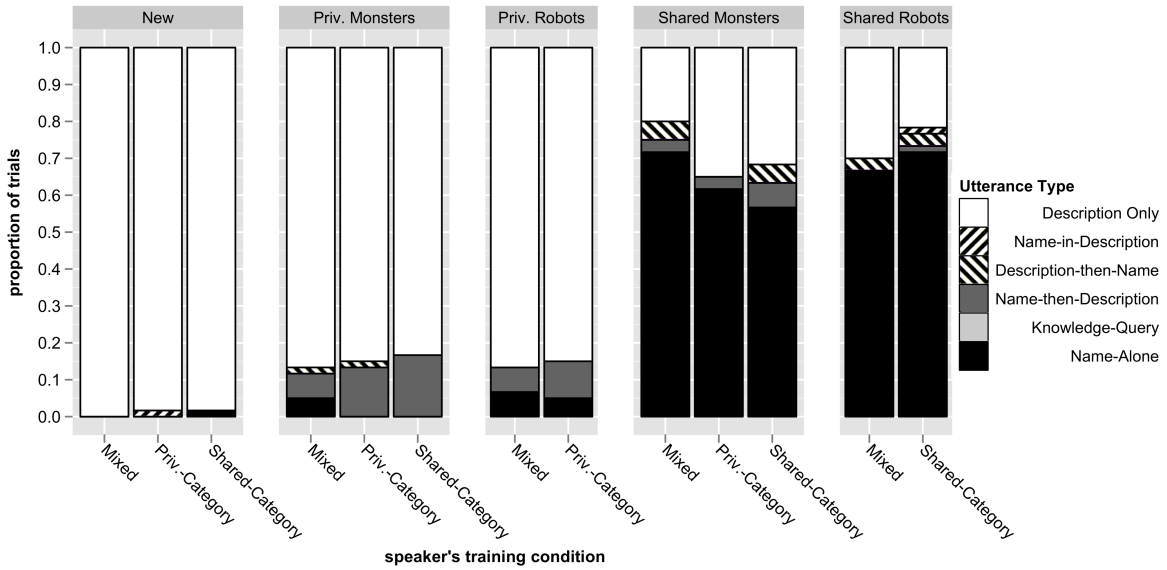

Figure 2. Experiment 1: Each panel shows the distribution of directors' utterances for different types of items (new, privileged monster/robots & shared monsters/robots). The distribution of the categories (monsters & robots) varied based on the training conditions (mixed, privileged-category & shared-category).

this form to mark information that they believe to be shared.

For statistical analyses, we collapsed the shared-category and privileged-category conditions and used a multilevel logistic regression model to predict whether directors used Name-Alone with the presence of category cues (category; mixed), the ground status of the test item (shared; privileged), the category type of the test item (robot; monster) and their two-way interactions as fixed effects and subjects and items as random intercepts. The three-way interaction could not included in the model because the data for privileged items was too sparse.

Each panel of Figure 2 shows the distribution of the speakers' utterance types for a particular type of item in each condition (i.e. for each distribution of monsters and robots). Directors almost exclusively used the No-Name form for new items (far left panel), which is expected since they did not learn names for these items. Overall, directors distinguished between shared items (far right panels) and privileged items (the 2nd and 3rd panels) in the form of their referring expressions. Directors rarely used the Name-Alone form for privileged items and use it frequently for shared items regardless of whether the items were monsters or robots. In the model, there was a significant main effect of ground status (ß=2.60, *SE*=0.32, $p<0.001$) such that Name-Alone was used more for shared items.

Additionally, for shared monsters (the 4th panel), speakers used the Name-Alone form less in the two category conditions. Remember that this is when the distribution of robots could be used as a cue to their ground status but the monsters spanned both shared and privileged ground. In the model, this was reflected by a significant interaction of category cues and an item's category type (ß=-1.31, *SE*=0.61, $p<0.04$) such that Name-Alone is used more for monsters when there are no category cues, but this pattern is reversed for robots. The presence of category cues makes directors more likely to use the Name-Alone form for items when all of the items in that category are shared, but less likely to use the Name-Alone form when there is a category cue and only some of items in that category are shared, indicating that directors are sensitive to the category manipulations.

Finally, it is interesting to note the trend for shared monsters that directors used the Name-Alone form more in the privileged-category (57%) than the shared-category condition (62%). In the privileged-category condition two-thirds of the monsters are shared, whereas in the shared-category condition only one-third of the monsters are shared. This pattern is similar to the difference between the high and low overlap conditions in Wu & Keysar (2007) and Heller et al. (in press).

Discussion

Overall, speakers are adept at distinguishing between shared and privileged names when shared knowledge is established via shared experience. The category manipulation produced only modest changes in their behavior, namely, reduced use of shared names for items in a category that is both shared and privileged. This shift in behavior suggests that categorical knowledge may play some role in how speakers distinguish between shared and privileged knowledge.

Experiment 2

Experiment 2 was designed to investigate the role of partner-specific shared experience on speakers' ability to distinguish shared and privileged knowledge. In this experiment, directors and matchers learned the same items as participants in the mixed condition of Experiment 1, but training was conducted separately. The matcher was alone when learning shared items and the director was either alone

or with a different partner who was not the matcher; participants were still aware of which items both had learned and which only the director learned. In the condition where directors learned shared items with a third party, the shared learning experience corresponds to the common ground they have with the matcher but does not provide a speaker-specific memory association that could be used during the referential communication task.

Method

Participants Ten pairs and 10 trios of participants participated in Experiment 2. Six additional pairs/trios participated but were excluded from analysis due to experimenter error (3 pairs), equipment failure (1 pair) or failure to follow task instructions (1 pair & 1 trio).

Materials The materials were identical to those used for the mixed condition of Experiment 1.

Procedure The same procedure as in Experiment 1 was used, except that the participants learned shared items separately rather than sitting next to each other. In the alone condition, the participants were told that the director would learn the names for some of the items, then the matcher would learn the same items, and then the director would learn the names for the remaining items, which would not be subsequently taught to the matcher. After all of the blocks were taught, the participants reviewed the names separately. In the third-party condition, the procedure for directors and matchers was identical to the alone condition, except that the director learned the shared items with a third participant. This additional participant was only present during training of items that the matcher also learned; the directors were alone when they learned privileged items.

Results

Task performance Matchers' task performance was excellent (they clicked on the correct item on 99% of the trials), but there was a notable difference in speakers' post-test performance. Directors' accuracy on the first post-test (distinguishing between items both they and their partners know and those only they learned) was 91% for the alone condition, which was significantly less than the third-party condition (96%; t=2.47, p<0.03) and the mixed condition of Experiment 1, which we will call the matcher condition here (98%; t=4.78, p<0.001). However, director's accuracy on the second post-test (memory for the items' names) was 78% for the alone condition and 85% for the third-party condition, which do not significantly differ from the matcher condition (84%).

Directors' Utterances Each panel of Figure 4 shows the distribution of directors' utterances for a different type of item used in Experiment 2. The alone and third-party conditions are compared to the matcher condition (the mixed condition of Experiment 1). For shared items (the right panel), there was no difference in the use of the Name-Alone form between the training conditions. However, for privileged items directors' use of the Name-Alone form increased when they had less shared experience, suggesting that directors were prone to errors (i.e. using names when the matcher didn't know them) when the director and matcher's shared knowledge was established without direct shared experience.

This pattern was confirmed when we compared a multilevel logistic regression model that predicts the likelihood of the Name-Alone form with item (shared; privileged), training type (matcher, third party, alone) and their interaction as fixed effects and with items and subjects as random intercepts to a similar model that did not contain the interaction. This comparison indicated that the addition of the interaction was significant (χ^2 = 23.47, p<0.002). To confirm the linear trend, we also constructed another multilevel logistic regression model containing the interaction term, but included the matcher, third party and alone conditions as -1, 0, 1, respectively, rather than as categorical predictors. This analysis showed a significant interaction of training and item type (ß=-0.54, SE=0.16, p<0.001)

Another notable difference in directors' utterances when they don't have a shared learning experience with the matcher is that directors used the Knowledge-Query form (e.g. *Do you know Inta?*), which further suggests that they were less certain about ground status of privileged items in the alone and third-party conditions than they were in the matcher condition.

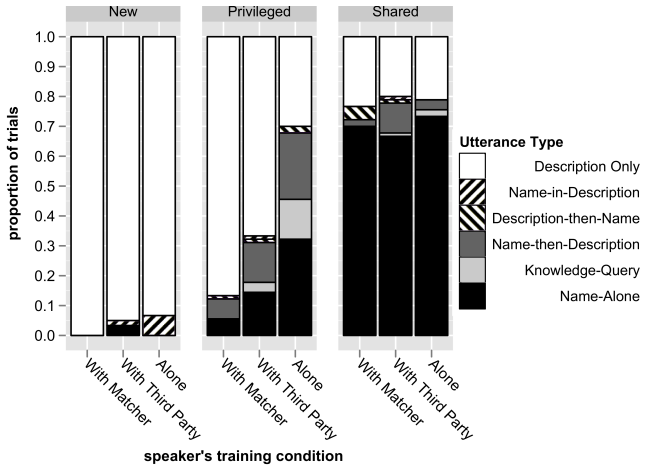

Figure 4. Experiment 2: Each panel shows the distribution of directors' utterances for new, privileged & shared items for each training condition (alone, third party & matcher).

Discussion

When directors learn the same information as their partner without experiencing the learning process together, even when they have another person who can act as a cue to their partner's knowledge, directors seem to have difficulty distinguishing between shared and privileged names, as shown in their explicit judgments in post-tests as well as their utterances. Although they use the Name-Alone form

equally often for shared items in all conditions, directors they use the Name-Alone form more for privileged items when there was not shared experience with the matcher.

General Discussion

These experiments demonstrate a remarkable ability in speakers for tracking the common ground status of a particular name with respect to a particular addressee, and shed some light on the nature of the memory representations that support that ability. When shared knowledge is established through shared experience, speakers are quite adept at tracking the status of names and using them appropriately in a referential communication task. This lends support to Horton and Gerrig's (2005a) memory-based hypothesis. Shared learning experience should create a strong episodic memory cue linking the partners to names, thus encoding ground status.

When the learning experience is instead shared with a third-party whose knowledge perfectly corresponds with the addressee, speakers are less certain about ground status, as demonstrated by their use of names for privileged items and explicit knowledge queries. This suggests that while this type of indirect cue can help speakers to distinguish between shared and privileged knowledge, it is not as helpful as having a shared learning experience with a conversational partner. However, it is more helpful than having no shared experience at all, which suggests that the speaker's confidence about ground status varies based on the quality of the evidence linking the addressee to the name of an item; that is, speakers appear to be relying on representations that are more graded than a "one-bit" model of their partner would seem to allow.

Shared experience clearly played the largest role in determining the ground status of a name. However, we also found that category structure can serve as a cue to ground status. Speakers were more likely to use the Name-Alone form in situations when all of the members of a category were learned together, and are less likely to do so when some members of a category were learned together and others were not. This, along with the third party condition in Experiment 2, suggests that speakers are also able to use sources of information about their partner's knowledge that are not based on having shared learning experience directly with them. In everyday conversations, we are likely to have more in common ground with our interlocutor than could possibly have been established strictly through shared experience, and as such, sources of information that do not depend on shared experience, such as category-based inferences about what an interlocutor knows, may become more important in those settings.

Horton and Gerrig (2005a) provide a useful model for how memory representations could underlie a speakers' ability to use felicitous referring expressions. Further study will need to address the question of how these memory processes scale up to more realistic kinds of world knowledge in more realistic conversational settings and how speakers use cues like shared experience and categorical structure under these conditions.

Acknowledgments

We are grateful to Dana Subik for recruiting participants and providing technical support. This research was partially supported by NIH grant HD-27206 to MKT and an NSF Graduate Research Fellowship to KSG.

References

Brennan, S.E. (2000). Processes that shape conversation and their implications for computational linguistics. In *Proceedings of the 38th Annual Meeting of the ACL*. Hong Kong: Association for Computational Linguistics.

Brennan, S.E., & Hanna, J.E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science, 1*(2), 274-291.

Clark, H.H., & Marshall, C. (1978). Reference Diaries. In D.L. Waltz (Ed.), *Theoretical issues in natural language processing, TINLAP-2* (pp. 57-63), ACM: New York.

Clark, H.H., & Schober, M. (1992). Understanding by Addressees and Overhearers. In Clark, H.H., ed., *Arenas of Language Use*, Chapter 6. University of Chicago Press.

Clark, H.H., & Wilkes-Gibbs, D. (1992). Referring as a collaborative process. In Clark, H.H., ed., *Arenas of Language Use*, Chapter 1. University of Chicago Press.

Creel, S.C., Aslin, R.N., & Tanenhaus, M.K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*, 633-664.

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

Heller, D., Gorman, K.S., and Tanenhaus, M.K. (in press). To name or to describe: shared knowledge affects referential form. *Topics in Cognitive Science.*

Horton, W.S. (2007). The influence of partner-specific memory associations on language production: Evidence from picture naming. *Language and Cognitive Processes*, *22*, 1114-1139

Horton, W.S., & Gerrig, R. (2005a). Conversational Common Ground and Memory Processes in Language Production. *Discourse Processes, 40*(1), 1-35.

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

Keysar, B., Lin, S. & Barr, D.J. (2003) Limits on theory of mind use in adults. *Cognition, 89(1),* 25-41

Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Metzing, C., & Brennan, S. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language, 49*(2), 201-213.

Searle, J. (1950). Proper Names. In *Mind, 59,* 320-344.

Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science, 31*, 1-13.