



Published in final edited form as:

Top Cogn Sci. 2012 April ; 4(2): . doi:10.1111/j.1756-8765.2012.01182.x.

To Name or to Describe: Shared Knowledge Affects Referential Form

Daphna Heller^a, Kristen S. Gorman^b, and Michael K. Tanenhaus^b

^aDepartment of Linguistics, University of Toronto

^bDepartment of Brain and Cognitive Sciences, University of Rochester

Abstract

The notion of common ground is important for the production of referring expressions: In order for a referring expression to be felicitous, it has to be based on shared information. But determining what information is shared and what information is privileged may require gathering information from multiple sources, and constantly coordinating and updating them, which might be computationally too intensive to affect the earliest moments of production. Previous work has found that speakers produce overinformative referring expressions, which include privileged names, violating Grice's Maxims, and concluded that this is because they do not mark the distinction between shared and privileged information. We demonstrate that speakers are in fact quite effective in marking this distinction in the form of their utterances. Nonetheless, under certain circumstances, speakers choose to overspecify privileged names.

Keywords

Common ground; Language production; Perspective taking; Referring expressions; Names

1. Introduction

When producing a referring expression, speakers have a wide range of options. For example, in referring to a particular New York City landmark, speakers can choose between the name in (1a) and the description in (1b).

(1) a. Rockefeller Center.

b. The square with all the international flags.

Understanding how and when speakers use a name or a description is important for developing and evaluating models of conversation. For example, it is likely to provide insights into audience design; in particular, under what circumstances and to what degree speakers tailor utterances for addressees. It is also important for developing conversational agents for task-based collaborative dialogs (Allen et al., 2001). As reference generation algorithms become more natural, it is likely to reduce the cognitive load associated with using a conversational agent (Campana, Tanenhaus, Allen, & Remington, 2011).

One factor that likely affects the choice between a name and a description is the assumed knowledge of the addressee, or what is in “common ground” (e.g., Clark & Marshall, 1981; Isaacs & Clark, 1987; Stalnaker, 1978). When directing her addressee’s attention to a landmark in a photograph, a speaker would be more likely to use (1a) rather than (1b) if she assumed that he was familiar with central landmarks in New York City. If the speaker uses a name that the addressee does not know, the referring expression will not allow the addressee to identify the intended referent, and further utterances will be required to establish reference. Therefore, if the name is not assumed to be known to the addressee, a cooperative speaker should use a description. If, however, the name *can* be assumed to be known to the addressee, the speaker in principle can choose between a name and a description. But a description takes longer to produce and thus is less efficient than a name.

According to Grice’s Maxims (Grice, 1975), a cooperative speaker should use a name, because a name is usually shorter than a description. In a classic paper, Dale and Reiter (1995) proposed an incremental computational model that realizes Grice’s Maxims while incorporating some basic assumptions about human lexical preferences. The model focuses on expressions that are generated for the purpose of enabling the addressee to identify the intended referent, assuming that optimal reference production involves selecting the shortest description based on salient properties that will allow the intended referent to be identified. However, as we will discuss shortly, psycholinguistic data suggest that speakers are not always Gricean.

If the form of referring expressions depends on common ground, this raises the general question of how a speaker determines what information is shared, that is, what is in common ground, and what information is privileged, that is, known to the speaker but not to her addressee. Distinguishing shared and privileged information may require gathering and coordinating indirect evidence from multiple sources, including the physical environment, the linguistic discourse and general information about the conversational partner, determining how reliable the evidence is, and updating the status of information as shared or privileged as new information becomes available in conversation. Because of the complexity that could be involved in storing, maintaining, and accessing common ground during conversation, it has been suggested that these computations are, in fact, too slow to affect the processing in real time. Keysar and colleagues have proposed that interlocutors do not take into account shared information in processing, which amounts to treating all information similarly and processing relative to the interlocutor’s egocentric perspective (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003).

For comprehension, there is a growing body of psycholinguistic evidence suggesting that, at least under certain circumstances, listeners can effectively distinguish shared and privileged information and use this distinction from the earliest moments of processing (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008; Brown-Schmidt, 2009a,b; Hanna, Tanenhaus, & Trueswell, 2003; Heller, Grodner, & Tanenhaus, 2008; Nadig & Sedivy, 2002; Wu & Keysar, 2007a; but cf. Barr, 2008). For production, however, using the distinction between shared and privileged information in planning might be a harder task. One reason is that speakers need to rely on their memory representations alone, as they do not get explicit cues from their conversational partner like those available to listeners interpreting the linguistic signal produced by the speaker. In addition, speakers cannot simply focus on shared information, because a felicitous assertion should contribute new information to common ground, forcing them to keep track of the shared versus privileged distinction. Finally, speakers have to balance the cognitive demands of formulating and producing utterances against the potentially resource-demanding task of determining what knowledge is shared and what is privileged.

Some existing studies suggest that speakers have difficulty in using the shared versus privileged distinction. Wardlow Lane, Groisman, and Ferreira (2006) examined the production of referring expressions where the intended referent had an object contrasting in size in privileged ground, and found that speakers sometimes produced a modified referring expression even though this kind of modification was inappropriate from the addressee's perspective, and were more likely to produce those modified expressions when they were explicitly told to conceal the identity of the privileged object from their partner. This pattern of results has led Wardlow Lane et al. to conclude that speakers are unable to inhibit irrelevant privileged information when producing referring expressions, and they produce using their egocentric perspective.

Wu and Keysar (2007b) proposed a plausible heuristic that speakers might use to determine which names are likely to be in common ground, without consulting specific information about ground. In particular, they suggest that speakers rely on the global information overlap with their addressees. In particular, their Information Overlap Heuristic states that "when overlap in information between two people is extensive, using one's own information should work just fine because it is most likely to be shared" (p. 4). They tested the overlap heuristic hypothesis in an experiment in which speakers learned artificial names for novel shapes, some together with their addressee and some alone. Participants then performed a referential communication task (Fussell & Krauss, 1989; Glucksberg, Krauss, & Weisberg, 1966) in which the speaker was presented with a target shape and she instructed the addressee about which of the shapes to click on. Wu and Keysar (2007b) assumed that speakers will follow Grice's Maxims as described for example (1), using names for shared shapes; using a description in this case would be under-informative and thus uncooperative. For shapes whose name is only known to the speaker, a name should not be used, because a privileged name will not be informative to the addressee, and is therefore likely to result in failed reference. Such optimal behavior should create a one-to-one correlation between name use and shared names, but Wu and Keysar found that speakers sometimes used privileged names that were not known to their addressees. Assuming that speakers would not intentionally violate Grice's Maxims, Wu and Keysar concluded that speakers were not able to keep track of the shared versus privileged status of names. Instead, because they found that speakers used significantly more names when they shared more information with their addressee, they concluded that speakers rely on a global heuristic.

There is evidence, however, that speakers are not always optimally informative as predicted by Grice's Maxims, providing more information than is strictly necessary to identify a referent (Engelhardt, Bailey, & Ferreira, 2006; Isaacs & Clark, 1987). For example, Isaacs and Clark (1987) had experts and novices on NYC landmarks perform a referential communication task ordering NYC postcards. Isaacs and Clark found that experts sometimes both named and described the landmark, as in (2).

(2) Rockefeller Center, with all the flags.

The assumed knowledge of the addressee for an utterance of this form is less straightforward than for (1a) or (1b). Isaacs and Clark found this form more often when experts were giving instructions to novices than to other experts, which suggests that speakers used this form when they did not assume that their addressee knew the name. But, in this study, there was no direct measure of whether participants knew specific names; participants were recruited based on their overall experience with NYC. Moreover, the speaker might have first uttered just the name, and only provided the description after she either received no feedback from her addressee or feedback indicating that he was confused. This would suggest that the speaker assumed that the name was shared. However, if the name and the description were

planned together, that would suggest that the speaker assumed that the addressee did not know the name, but chose to introduce a new name for a future use. Importantly, for Wu and Keysar, both cases would be taken to indicate that the speaker assumed that the name was shared. With this in mind, we replicated Wu and Keysar's (2007b) experiment but looked in greater depth at the forms used by speakers in order to determine whether name use is similar for shared and privileged objects.

2. Methods

2.1. Participants

Forty pairs of naïve participants are included in the analysis. They were all native speakers of English recruited from the University of Rochester community and were paid \$15 each for their participation.

Participants knew each other beforehand and chose to participate in the experiment together. One pair was excluded from the analysis because they were unable to complete the training and another because of computer failure. Six additional pairs were excluded because they misunderstood the instructions.¹

2.2. Materials and procedure

The 24 novel shapes and their artificial names were adapted from Wu and Keysar (2007b) such that 6 of the 24 names shared the onset /fl/.² Thirty additional novel shapes were used as distractors during the testing phase only.

2.2.1. The training phase—The two participants sat together across from the experimenter, who had index cards with the 24 named shapes. Participants learned the names of the shapes in four blocks of six. On each trial, the experimenter presented a card, articulated the name, and waited for both participants to repeat the name before proceeding to the next card. After going through the six shapes in the block once, the experimenter presented the card and waited for the participants to name the shape; the experimenter articulated the name or corrected any errors if the participants could not name the shape correctly. The experimenter repeated this procedure for the block until both participants could name all six shapes flawlessly, and then moved on to the next block. After all the blocks had been learned, the experimenter had the participants name the shapes in each block before proceeding to the testing phase.

Common ground was manipulated by having the (randomly selected) matcher learn only a subset of the names learned by the director. Participants first learned some names together; then the director continued to learn more names alone, and the matcher played a

¹We had initially excluded these pairs without examining their data. At the suggestion of the editors and reviewers, we looked at the data from the six pairs to determine whether excluding these pairs might be reducing evidence for the overlap heuristic. The directors in all of these pairs did not seem to understand that their task was to give instructions to their partner. Two pairs asked about this midway through the experiment. The directors for another three pairs said, "I don't know" or "I don't know that one" whenever they did not know the name of the shape without making any attempt to tell their partner what to click on. These pairs also ignored the error signal when the matcher made a mistake. The sixth pair made frequent errors, the matcher never spoke, and the pair ignored the error signal.

²The current study was originally designed to build upon Wu and Keysar's results by manipulating not just Overlap but also Category: the structure of information. Category, which like Overlap, was a between-subjects factor, manipulated whether the six shapes in the last block of training shared the onset /fl/ (Category conditions) or had no properties in common (No-Category conditions). Our hypothesis was that perceptual grouping of the stimuli would allow speakers to better track the distinction between shared and privileged information. However, when we analyzed the data, it became clear that speakers in the No-Category conditions were already at ceiling in distinguishing shared and privileged information, making the Category manipulation irrelevant. Indeed, there were no effects or interactions of Category. For ease of exposition, at the suggestion of reviewers, we have removed discussion of the category factor from the body of the article. It is important to note that the same shapes were used in the Category and No-Category conditions for the referential communication task, so these are equivalent to two lists.

nonlinguistic computer game while listening to music over headphones (the matcher stayed in the same room). Following Wu and Keysar (2007b), we manipulated the relative amount of information in common versus privileged ground. In the High Overlap conditions, participants learned 18 names together, whereas in the Low Overlap conditions participants learned only 6 names together.

2.2.2. The testing phase: A referential communication task—Participants sat in front of two different computers in different rooms and were free to converse over a network. The director was presented with one shape (Shared, Privileged, New) and had to instruct the matcher who saw three shapes, to click on the target shape “as quickly and accurately as possible.” The matcher’s display contained one shape with a shared name, one shape with a name that was privileged to the director (i.e., unknown to the matcher himself), and one new shape that was unnamed: The two distractors were randomly chosen from the relevant set of shapes. Trials were advanced when the matcher clicked on any shape: If the matcher clicked the wrong shape, an error sound was heard, but participants could not correct the error. The referential communication task had two practice trials followed by 18 experimental trials, six of each shape type (Shared, Privileged, New).

The 12 shared and privileged shapes used in testing were identical across all training conditions. The shared shapes were drawn from the first block of training, and the privileged shapes were drawn from the fourth (i.e., last) block of training that was privileged to the director in both levels of Overlap. As a result, testing always included the six /fl/ names and six other names (see again Footnote 2). This resulted in a 2×3 design, crossing Overlap (High vs. Low: between-subjects factor) with Type of Shape, hereafter referred to as Ground (Shared, Privileged, New shapes: within-subject factor).

2.2.3. Posttests and debriefing—After completing the referential communication task, the speaker had to complete two additional tasks. First, the director was presented with the 24 shapes she had learned during training and had to determine for each shape whether she had learned the name of that shape together with her partner or alone. The computer presented the shapes one at a time in random order, and the director had to click “learned together” or “learned alone.” After that, the director completed a second task where she had to name each of the shapes. The computer presented the shapes one at a time in (a different) random order, and the director had to say its name, which was recorded onto the computer.

Once the posttesting was completed, both participants were debriefed. The director was first debriefed alone, and the matcher was debriefed in the presence of the director. We first asked general questions about the experiment, and later continued to ask specifically about any strategies the director might have used in naming the shapes.

3. Results

3.1. Task performance

Task performance was excellent. In the referential communication task, matchers clicked on the correct shape on 98% of the trials. In the first posttest, which was to distinguish between shapes that were learned together or alone, accuracy was 95%, and there was no significant difference between accuracy in the High and Low Overlap (96% vs. 93%). In the second posttest, which was to name 24 shapes, overall performance was 83% with no significant difference between the High and Low Overlap (86% vs. 80%). Thus, speakers were remarkably good at remembering which names were shared and which were privileged.

3.2. The form of the utterances

We assigned each of the speaker's utterances to one of five categories: name alone, name followed by description, description followed by a name, description containing a name (where this name was not the name of the target), and description without a name. Examples of these utterance types are presented in Table 1. Fig. 1 presents the distribution of the five utterance types for the six conditions in the referential communication task. This figure reveals clear effects of both Overlap (High vs. Low) and Ground (Shared, Privileged, New). When Overlap was Low, there were more descriptions and fewer names than when Overlap was High. More important, in the Shared conditions most utterances included names, whereas in Privileged and New conditions most utterances included descriptions, although more names were used in Privileged conditions than in New conditions. Most strikingly, the form of utterances with names is strongly affected by Ground. In Shared conditions, speakers typically used the name alone, whereas in Privileged conditions, the name is almost always followed by a description. We analyze this pattern after examining the data according to Wu and Keysar's (2007b) coding categories.

3.2.1. Wu and Keysar's coding categories—When we collapsed across utterance types to create the categories used in the two analyses reported by Wu and Keysar, both the data pattern and the proportion of utterances in each category were nearly identical to those reported in Wu and Keysar (2007b). In their first analysis, utterances were classified according to whether a name was used anywhere in the referential description; we will refer to this analysis as the “all names” analysis. This corresponds to the three utterance types that include the name of the target shape (name-alone, name-then-description, description-then-name): Shared (High 0.78 vs. Low 0.65), Privileged (0.34 vs. 0.16), and New (0.05 vs. 0). Following Wu and Keysar, proportions were submitted to a 2 (Overlap: between-subjects factor) \times 3 (Ground: within-subjects factor) ANOVA (all proportions were logit-transformed before ANOVA; see Jaeger, 2008, on the benefits of the logit transformation as compared with the arcsine transformation). There was a main effect of Overlap, $F(1,38) = 9.09$, $p < .01$, as well as a main effect of Ground, $F(2,76) = 91.60$, $p < .001$. Like Wu and Keysar, our data showed that speakers in the High Overlap training condition uttered significantly more names overall for Privileged shapes than speakers in the Low Overlap training condition, $F(1,38) = 4.42$, $p < .05$.

Wu and Keysar also performed a second analysis, including just those utterances in which the name occurred before any description of the shape. The goal of this analysis was to exclude cases “in which speakers first described the object in order to identify it and then named it in order to inform their addressee about the name” (Wu & Keysar, 2007b, p. 7). This corresponds to our name-alone and name-then-description categories: Shared (High 0.70 vs. Low 0.51), Privileged (High 0.29 vs. Low 0.12), and New (High 0.05 vs. Low 0). Like Wu and Keysar, we found that excluding description-then-name trials did not change the pattern of results. A 2 \times 3 ANOVA performed on logit-transformed proportions revealed main effects of both Overlap, $F(1,38) = 12.36$, $p = .001$, and Ground, $F(2,76) = 86.15$, $p < .001$. Again, High Overlap directors were significantly more likely to use names in referring to Privileged shapes than their Low Overlap counterparts, $F(1,38) = 5.65$, $p < .05$.

Both analyses indicate that names were more likely to be used by High Overlap directors than by Low Overlap directors overall, but crucially this was also the case for Privileged shapes. On the basis of these data, Wu and Keysar concluded that overlap is used as a heuristic that substitutes knowledge of ground for individual items. However, it is important to note that ground was a significant factor in determining whether speakers used a name. Most important, the distribution of utterance type within the name categories also calls into

question the assumption that a speaker who uses a privileged name is unaware that her interlocutor does not know that name.

3.2.2. A four-way distinction in name use—The most striking difference between how names are used for shared and privileged shapes was that most of the utterances with names in Shared conditions were name-alone utterances, whereas name-alone utterances were rarely used in the Privileged conditions: Shared (High 0.64 vs. Low 0.48), Privileged (0.05 vs. 0.01), and New (0.05 vs. 0). Proportions were logit-transformed and submitted to a 2 (Overlap) \times 3 (Ground) ANOVA. Here, too, there was a main effect of Overlap, $F(1,38) = 7.11$, $p < .05$, but, unlike in Wu and Keysar's analyses, the difference was not significant for Privileged shapes, $F(1,38) = 1.26$, $p = .27$. There was also a main effect of Ground, $F(2,76) = 219.25$, $p < .001$. The name-alone form was more often used in Shared conditions than in Privileged conditions, $F(1,39) = 328.04$, $p < .001$, but Privileged and New did not differ ($F < 1$). Thus, when we focused on an utterance type that does not include any descriptive content that could potentially help the matcher identify the intended referent, we no longer found any evidence that directors were unaware of the privileged status of shapes and were using a global strategy to determine their status. Instead, we observed the expected behavior from a cooperative speaker where names are used almost exclusively for shapes with shared names.

In the Privileged conditions, directors mostly used a name followed by a description, as in (2) above; this utterance type was rarely used in Shared conditions, and never in New conditions: Shared (High 0.06 vs. Low 0.03), Privileged (High 0.24 vs. Low 0.11), and New (0 vs. 0). The main effect of Overlap was found here as well, $F(1,38) = 5.67$, $p < .05$, and here it was significant when comparing the two levels of Privileged, $F(2,38) = 6.38$, $p < .05$. Again, there was also a main effect of Ground, $F(2,76) = 19.73$, $p < .001$, with the difference between Privileged and Shared conditions being significant, $F(1,39) = 9.97$, $p < .01$.

In sum, as observed by Wu and Keysar (2007b), we found that directors indeed used more names for privileged shapes than for new shapes, and that name use increased with overlap. However, examining a wider range of utterance types revealed that directors use different forms for shared and privileged names. This suggests that although speakers use privileged names when name use is unexpected by Grice's Maxims, they are nonetheless aware of the privileged status of those names.

3.3. Listeners' sensitivity to forms

The analysis of the forms used by speakers reveals that they distinguish shared and privileged names in the form of their utterances, but in more subtle ways than would be expected from Grice's Maxims. It is therefore important to explore whether listeners are sensitive to the distinctions made by speakers. We report two preliminary analyses that examine whether the director's utterances provided sufficient information for listeners to determine whether the director assumed that the name was shared or privileged.

3.3.1. Global assessment—In order to test the idea that listeners are sensitive to these subtle distinctions between the utterance types used for shared and privileged names, we asked three naïve coders to listen to the conversations and judge for every name whether they thought the director assumed that the matcher knew that name. Note that although directors sometimes explicitly mentioned their assumptions about the status of the names, such explicit comments were found only for 3% of trials, so coders could not rely on these comments (see Examples 2.1, 3.3, and 3.4 in Table 1). Coders were also instructed to base their judgments solely on what the director said and *not* on the matcher's reaction to the director's utterance.

The coders were unaware of the goals of the experiment, the manipulations in training, and, most important, they were blind to the status of individual shapes. Coders listened to the conversations as a whole and were asked to (a) exclude utterances where names were not used or were used as part of a description; (b) exclude utterances where the director repeated a name first uttered by the matcher (because these are not informative with respect to the status of names for the director); (c) classify names as “assumed shared” if the coder thought that the director expected the matcher to know the name; and (d) classify names as “assumed privileged” if the coder thought that the director did not expect the matcher to know the name. In classifying utterances into these four categories, there was agreement among all three coders for 83% of trials.

Fig. 2 shows the distribution of trials that were judged as “assumed shared” and “assumed privileged” out of all trials. Of those trials that used names, we found that “assumed shared” judgments were overwhelmingly assigned to Shared trials, whereas “assumed privileged” judgments were mostly assigned to Privileged trials. For purposes of statistical analysis, we examined the proportion of trials that were judged as “assumed shared” out of all trials and submitted the logit-transformed proportions to a 2 (Overlap) \times 3 (Ground) ANOVA. The main effect of Overlap observed in the form analyses persisted here, $F(1,38) = 7.34$, $p = .01$, but crucially, the difference between High and Low Overlap for Privileged trials was not significant ($F < 1$). There was also a main effect of Ground, $F(2,76) = 195.29$, $p < .001$, because there were significantly more “assumed shared” names for Shared trials than for Privileged trials, $F(1,39) = 265.60$, $p < .001$, but “assumed shared” judgments were not more likely for Privileged trials than for New trials ($F < 1$). A parallel ANOVA for “assumed privileged” trials again revealed a main effect of Ground, $F(2,76) = 26.41$, $p < .001$, this time because “assumed privileged” judgments were more likely for Privileged trials than for Shared trials, $F(1,39) = 19.51$, $p < .001$. In this case, there were no significant effects of Overlap.

The results of the global judgment analysis provide preliminary evidence that listeners are sensitive to the way speakers marked the distinction between shared and privileged names. It seems likely that the two forms (“name-alone” and “name-then-description”) were the cues used by our coders in their judgment. This was supported by examining how the two correlate. Of the “name-alone” trials, 97% were judged as “assumed shared” and only 1% were judged as “assumed privileged.” Of the “name-then-description” trials, by contrast, only 14% were judged as “assumed shared” and 86% were judged as “assumed privileged.” This pattern suggests that the form of the utterance was a reliable cue for judging the status of names as shared or privileged. In future research, it will be important to determine whether listeners make use of these cues in real-time conversation.

3.3.2. “Name-then-description” as a repair strategy?—The name-then-description strategy that directors used for privileged names raises the question of whether this form was included in the earliest moments of utterance planning. Perhaps the director initially planned the name without taking ground into account and added the description as a kind of elaboration when she realized that she was uttering a name that was privileged to her and thus uninformative for the matcher. This type of postnominal after-thought or repair strategy occurs when new information arises late in the planning process. Crucially, it does not seem to affect how the initial name is uttered, for example, its fluency (Brown-Schmidt & Tanenhaus, 2006). Thus, if ground information was consulted after the name was planned, this process should not affect the form of the name. The question, then, is whether name-then-description is a repair strategy directors used when they detected an error while monitoring their own speech, or whether this is the form they chose early on when they were planning to refer to a shape with a privileged name.

Most trials categorized as name-then-description did not seem to be obvious repairs. Specifically, there were two kinds of clear repair: (i) the description was added after a long break (over two seconds) in which the matcher did not say anything and did not click on any shape; (ii) the director interrupted herself in the middle of the name and started producing a description. But in all of our data, there were only one trial of type (i) and two trials of type (ii)—see Example 2.4 in Table 1. Importantly, the vast majority of name-then-description trials did not seem to involve a break or an interrupted name.

We hypothesized that if directors were planning a name-then-description utterance from the earliest moments, this should be reflected in the way they pronounced the names, because the subsequent description would be in the same intonational phrase. This was expected to differ from name-alone trials, where the name occurs in the end of the intonational phrase. We tested this prediction with a new set of naïve listeners. We presented utterances that were truncated at the end of the name and asked listeners whether they expected the utterance to continue.

Twenty-four naïve listeners, who had not participated in the original study, listened to a set of truncated utterances from one original director and, after each name, judged whether they expected a continuation. Most directors uttered only a few names; therefore, we used the names from the four directors who produced five or more instances of name-alone and name-then-description: Director #21 (7 name-alone and 3 name-then-description), Director #22 (3 and 2), Director #30 (5 and 6), and Director #54 (4 and 5). Each participant listened to utterances from one director (six participants per director) and judged on a seven-point scale how likely a continuation was: 1 was characterized as “definitely nothing after”; 7 was characterized as “definitely something after.” Participants received written instructions for this task and had no exposure to the director’s speech apart from hearing the truncated utterances.

Listeners assigned a significantly higher rating to names that had continuations (mean score: 4.31) and those that did not (mean score: 3.05). We analyzed the data using a mixed-effects linear regression model with one fixed effect (the utterance type from which the name was extracted) and random effects for the slope and intercept for both subjects and items as well as a random intercept for director identity (i.e., the original director in the referential communication task). Ratings were significantly correlated with utterance type ($\beta = 1.25$, $p < .05$).

The difference in ratings for the two types of names, drawn from different types of utterances, provides preliminary evidence that directors pronounced names differently depending on whether they included a subsequent description in their utterance. This result suggests that the description that followed the name was planned early rather than being appended once the speaker realized that she was using an uninformative name. Therefore, speakers accessed ground information early in utterance planning.³

4. General discussion

We replicated Wu and Keysar’s (2007b) finding that speakers are more likely to use names when more information is shared. However, we demonstrated that this is not because they

³A reviewer suggested that the timing of utterance onset might also provide information about whether the name and description were planned together. If the description had been planned with the name, then the planning time before the onset of the speaker’s name-then-description utterance should be more similar to the planning time for description-alone trials than the planning time for name-alone trials. As a preliminary check, we measured the utterance-onset-times in milliseconds for the name-alone (1,147 ms), description (1,272 ms), and name-then description (1,347 ms) conditions. The data are too sparse for statistical analysis, but the pattern is consistent with our conclusion that the name and description were planned together.

use the global information overlap to estimate the shared or privileged status of individual names. Instead, we found that speakers clearly distinguished shared and privileged names in the form of their utterances, and that listeners seem to be sensitive to this distinction, even if they only listen to the very beginning of the utterance. When a speaker used a privileged name, it was not because she was unable to distinguish shared from privileged information, but rather because she chose to overspecify the referring expression, including a privileged name in addition to the description. These results complement previous studies showing that speakers do not always strictly follow Grice's Maxims, but sometimes choose to produce overspecified referring expression to meet other conversational goals (Engelhardt et al., 2006; Isaacs & Clark, 1987).

Why, then, did speakers chose to use names for privileged shapes and why were they more likely to use names in High Overlap condition compared with the Low Overlap condition? One possibility, which we are currently exploring, is that speakers were trying to teach the names to their addressees, hoping that it would make communication more effective if the shape occurred again during testing. Although every shape occurred only once, speakers did not know this before the referential communication task actually ended. In fact, during debriefing, a number of our participants volunteered that they used this teaching strategy. Teaching the addressee the names for the privileged shapes would make more sense when the addressee already knows most of the names. This strategy might not be specific to the circumstances in this experiment. Introducing names in a situation where there is some overlap in knowledge might be a generally efficient strategy. Thus, Wu and Keysar (2007b) might be correct in their hypothesis that speakers use an overlap strategy with respect to names. However, this overlap strategy is not used in lieu of information about whether a name is shared.

Our results demonstrate that speakers reliably distinguish between shared and privileged information, at least under conditions where they have first encountered that information in a shared context. They do not, however, adjudicate between two possible mechanisms for how interlocutors keep track of which names are shared. In a seminal paper, Clark and Marshall (1981) proposed that participants in a conversation create a specialized memory structure that stores partner-specific information, and that speakers consult these "reference diaries" in order to determine whether information is shared or privileged. However, Horton and Gerrig (2005a,b) argue that speakers do not need to construct complex representations such as reference diaries. Rather, they propose that knowledge about whether information is shared is an emergent property of ordinary memory processes. In particular, information about the conversational partner is encoded as part of domain-general episodic traces, and this person acts as a highly effective cue for the retrieval of relevant knowledge. This account seems plausible for the current results because there is evidence that lexical representations contain indexical information (Goldinger, 1996) and that indexical information is rapidly accessed in word recognition (Creel, Aslin, & Tanenhaus, 2008), although less is known about accessing indexical information in production. Because speakers first encountered the shapes and names in a shared learning experience, information about ground is likely to be encoded in the context of the name. What is striking about our results, then, is not that the information about ground was available in memory, but rather that speakers consider it early on in utterance planning.

Most knowledge that is likely to be shared between interlocutors is not of course knowledge that they first acquired together. Moreover, there is an extensive literature demonstrating that source memory is often inaccurate (Johnson, Hashtroudi, & Lindsay, 1991). However, as a conversation unfolds, interlocutors are likely to provide each other with information about what they do and do not know. An important question for future research is how interlocutors in a conversation update their assumptions about the likelihood that a specific

piece of knowledge is shared or privileged (e.g., whether an interlocutor is likely to know NYC landmarks) on the basis of feedback from the addressee, and how this updated information is reflected in the form of their referential expressions. As a first approximation, this could be modeled within a Bayesian framework as a process whereby interlocutors update their priors about the likelihood that their interlocutor will be familiar with a name. Sensitivity to what knowledge is shared and what is not would be reflected in the extent to which (a) interlocutors provide information about both their knowledge and the assumed common ground in the form of their utterances and (b) interlocutors attend to this information and modify utterances appropriately.

Our results have clear implications for developing conversational agents for collaborative tasks in that they provide psycholinguistic evidence about human behavior when choosing between names and descriptions. Specifically, the results show that speakers are not strictly Gricean in their referring expressions. In particular, they sometimes will use a name even when it is not assumed to be shared with the addressee, and they are more likely to do so when there is greater overlap between the knowledge of the interlocutors. A system that aims for naturalness will need to incorporate such overspecified descriptions in addition to names and descriptions. This is parallel to Dale and Reiter's (1995) point that systems should be able to produce referring expressions with overspecified color adjectives. Interestingly, in both cases, the overspecified information precedes the necessary information (Color + Noun or Name + Description), suggesting that overspecification may arise from incremental utterance planning. Unlike color adjectives that depend on perceptual information that can be assumed by a speaker (or a conversational agent) to be known to the addressee, assessing whether a specific addressee knows a certain name depends on nonperceptual information. The line of research we are proposing will provide evidence about how humans assess the status of information and update their estimates based on the referential choice of their partners. Similarly, a dialog system may also use the form of the users' referential choices to draw inferences about what the user assumes it knows and provide appropriate feedback.

Acknowledgments

We are extremely grateful to Dana Subik for recruiting, training, and testing all of the participants. We also thank Laura Zimmerman, Danielle Abrams, and Andrew Wood for help with data coding. This research was partially supported by NIH Grant HD-27206 to MKT and an NSF Graduate Research Fellowship to KSG.

References

- Allen J, Byron DK, Dzikovska M, Ferguson G, Galescu L, Stent A. Towards conversational human-computer interaction. *AI Magazine*. 2001; 22(4):27–37.
- Barr DJ. Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*. 2008; 109:18–40. [PubMed: 18760407]
- Brown-Schmidt S. The role of executive function in perspective-taking during on-line language comprehension. *Psychonomic Bulletin and Review*. 2009a; 16:893–900. [PubMed: 19815795]
- Brown-Schmidt S. Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*. 2009b; 61:171–190. [PubMed: 20161117]
- Brown-Schmidt S, Gunlogson C, Tanenhaus MK. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*. 2008; 107:1122–1134. [PubMed: 18164700]
- Brown-Schmidt S, Tanenhaus MK. Watching the eyes when talking about size: An investigation of message formulation and utterance planning. *Journal of Memory and Language*. 2006; 54:592–609.
- Campana E, Tanenhaus MK, Allen JF, Remington R. Natural discourse reference generation reduces cognitive load. *Journal of Natural Language Engineering*. 2011; 17(3):311–329.

- Clark, HH.; Marshall, CM. Definite reference and mutual knowledge. In: Joshi, AK.; Webber, BL.; Sag, IA., editors. *Elements of discourse understanding*. Cambridge, England: Cambridge University Press; 1981. p. 10-63.
- Creel SC, Aslin RN, Tanenhaus MK. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*. 2008; 106:633–664. [PubMed: 17507006]
- Dale R, Reiter E. Computational interpretations of the Gricean Maxims in the generation of referring expressions. *Cognitive Science*. 1995; 19(2):233–263.
- Engelhardt P, Bailey KGD, Ferreira F. Do speakers and listeners obey the Gricean Maxim of Quantity? *Journal of Memory and Language*. 2006; 54:554–573.
- Fussell SR, Krauss RM. Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*. 1989; 19:509–526.
- Glucksberg S, Krauss RM, Weisberg R. Referential communication in nursery school children: Method and some preliminary findings. *Journal of Experimental Child Psychology*. 1966; 3:333–342. [PubMed: 6006725]
- Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*. 1996; 22:1166–1183.
- Grice, H. Logic and conversation. In: Cole, P.; Morgan, J., editors. *Syntax and Semantics, Vol. 3, Speech Acts*. New York: Academic Press; 1975. p. 41-58.
- Hanna JE, Tanenhaus MK, Trueswell JC. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*. 2003; 49:43–61.
- Heller D, Grodner D, Tanenhaus MK. The role of perspective in identifying domains of reference. *Cognition*. 2008; 108:831–836. [PubMed: 18586232]
- Horton WS, Gerrig RJ. Conversational common ground and memory processes in language production. *Discourse Processes*. 2005a; 40(1):1–35.
- Horton WS, Gerrig RJ. The impact of memory demands upon audience design during language production. *Cognition*. 2005b; 96:127–142. [PubMed: 15925573]
- Isaacs EA, Clark HH. References in conversations between experts and novices. *Journal of Experimental Psychology: General*. 1987; 116:26–37.
- Jaeger TF. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*. 2008; 59:434–446. [PubMed: 19884961]
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin*. 1991; 114:3–28. [PubMed: 8346328]
- Keysar B, Barr DJ, Balin JA, Brauner JS. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*. 2000; 11:32–37. [PubMed: 11228840]
- Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. *Cognition*. 2003; 89:25–41. [PubMed: 12893123]
- Nadig AS, Sedivy JC. Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*. 2002; 13:329–336. [PubMed: 12137135]
- Stalnaker, R. Assertion. In: Cole, P., editor. *Syntax and semantics 9: Pragmatics*. New York: Academic Press; 1978. p. 315-332.
- Wardlow Lane L, Groisman M, Ferreira VS. Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science*. 2006; 17:273–277. [PubMed: 16623681]
- Wu S, Keysar B. The effect of culture on perspective taking. *Psychological Science*. 2007a; 18:600–606. [PubMed: 17614868]
- Wu S, Keysar B. The effect of information overlap on communication effectiveness. *Cognitive Science*. 2007b; 31:1–13. [PubMed: 21635286]

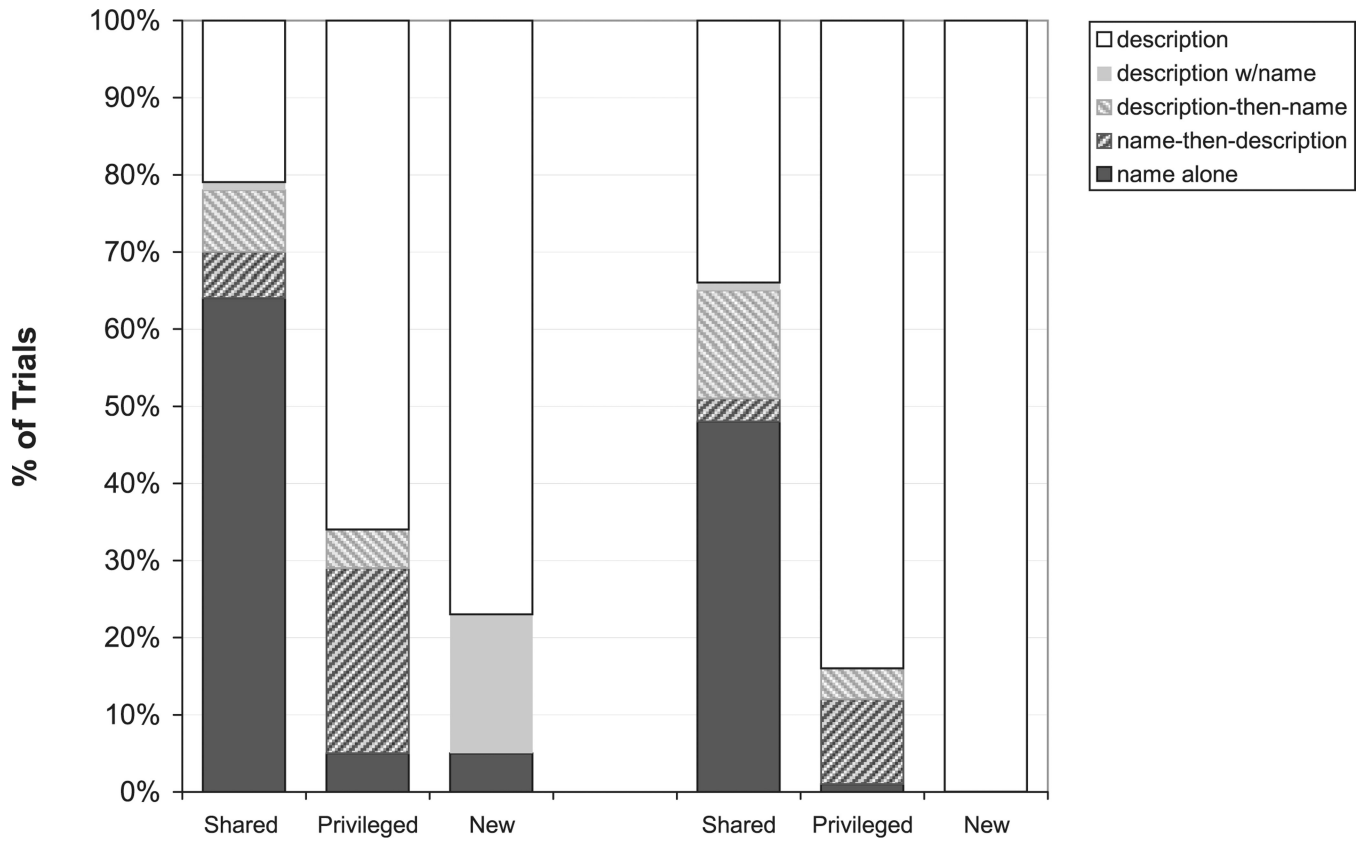


Fig. 1. The distribution of utterances across the five types for each of the six conditions.

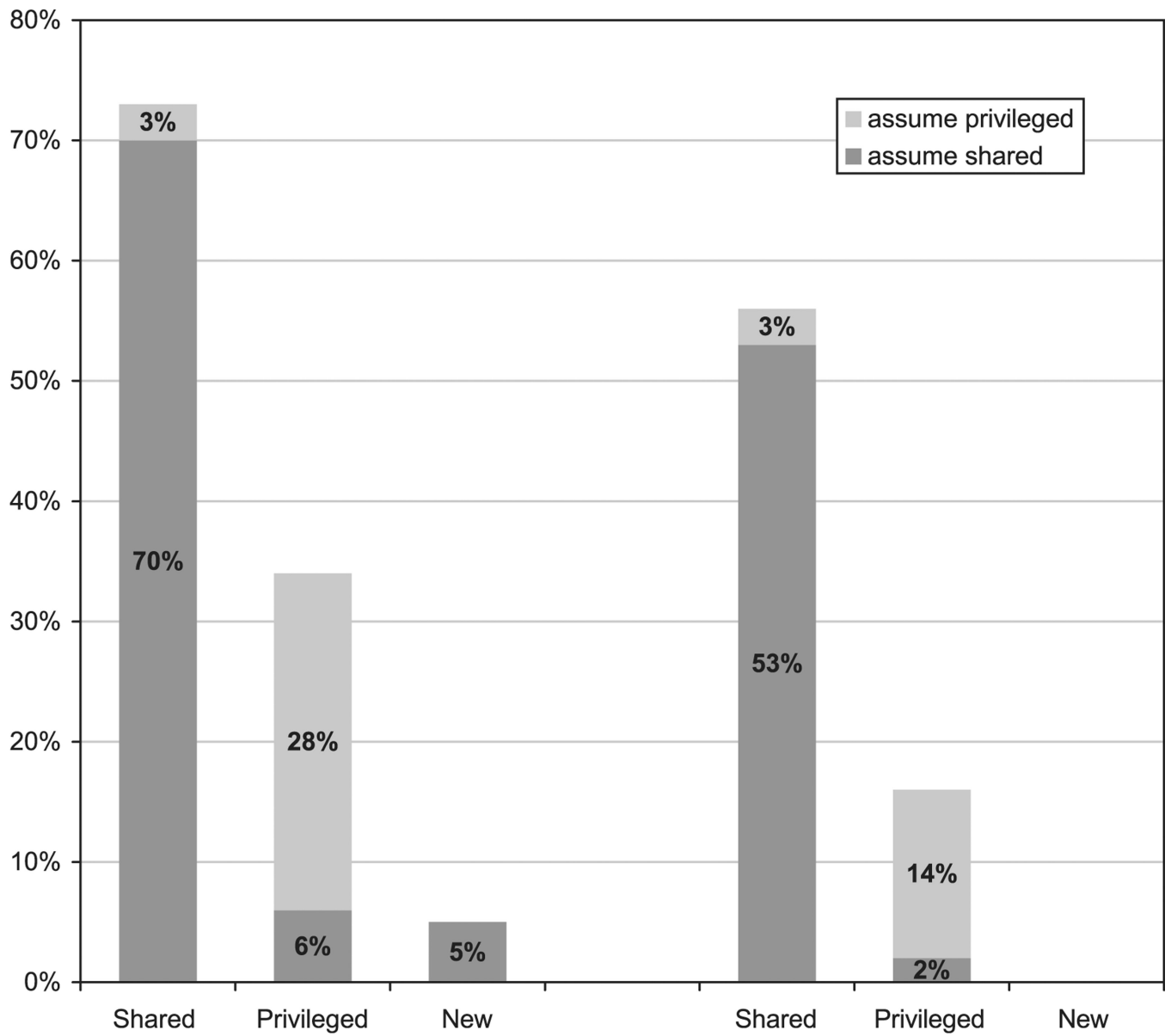


Fig. 2. Percentages of trials judged as “assume shared” and “assume privileged” out of all trials.

Table 1

Examples of utterances from the five categories

Name alone

- (1.1) Uhm cortlog.
 (1.2) Ah, banpar, your favorite.
 (1.3) Um, abypit I think it's called, I forget.
 (1.4) It's like another, it's like an abypit. It's just a ...yeah.

Name-then-description

- (2.1) Ah, inta, you haven't seen it, it's four arrows.
 (2.2) Uhm cortlog it's somebody hunching towards the left and pointing that direction.
 (2.3) Ah, this is called molget, it's like a triangle and a rectangle.
 (2.4) Um, flu- it's like, it looks like a person sort of.

Description-then-name

- (3.1) Looks like a rabbit, flanzo.
 (3.2) Ah, a box and triangle, it's a molget is the name of it.
 (3.3) It's a flag, you know that one I think, banpar.
 (3.4) Oh you don't know this, ok it's square with arrows coming out of it... it's called floogle if you were interested.

Description-with-name

- (4.1) Um, it's kind of like a cortlog but it's not.
 (4.2) This is like an etrett except on its side.
 (4.3) Ah, grampent, except weirdly faced.
 (4.4) Ah, it looks like the chicapee one except with a long tail.

Description

- (5.1) This one looks like the sun except the arrows are all together.
 (5.2) Um it's two triangles kissing.
 (5.3) It's got, it's a box with arrows coming out of the sides of it.
 (5.4) Uh, a person sitting down with their legs out.