Published in final edited form as:

Cognition. 2008 June ; 107(3): 1122–1134.

Addressees distinguish shared from private information when interpreting questions during interactive conversation

Sarah Brown-Schmidt,
University of Illinois at Urbana-Champaign

Christine Gunlogson, and
University of Rochester

Michael K. Tanenhaus
University of Rochester

Abstract

Two experiments examined the role of common ground in the production and on-line interpretation of wh-questions such as *What's above the cow with shoes?* Experiment 1 examined unscripted conversation, and found that speakers consistently use wh-questions to inquire about information known only to the addressee. Addressees were sensitive to this tendency, and quickly directed attention toward private entities when interpreting these questions. A second experiment replicated the interpretation findings in a more constrained setting. These results add to previous evidence that the common ground influences initial language processes, and suggests that the strength and polarity of common ground effects may depend on contributions of sentence type as well as the interactivity of the situation.

Keywords

common ground; eye-tracking; perspective taking; conversation; referential communication; question; comprehension

Conversation is a joint activity through which interlocutors exchange information and ideas. According to one influential framework, during conversation, interlocutors develop common ground (Stalnaker, 1978), which includes mutual knowledge that arises from common cultural and community information, information present in a shared environment, and information introduced linguistically (Clark, 1992, 1996).

Recently, psycholinguists have begun to explore when, in the time course of processing, interlocutors make use of common ground in interpreting definite referring expressions. Most of these studies have used the visual world eye-tracking paradigm (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995) to examine whether addressees look at referents that are potentially consistent with a referring expression, presented as an instruction, e.g., *Put the tape above the apple*, when that referent is in their privileged ground (e.g., a display including a

Contact Information: Sarah Brown-Schmidt, Beckman Institute, 405 N Mathews Ave, University of Illinois, Urbana-Champaign, Urbana, IL 61801, brownsch@uiuc.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cassette tape, an apple and several other objects visible to both participants, along with a roll of scotch tape that is occluded from the speaker). There is general agreement that the addressee's interpretation is restricted to candidates in common ground; however, studies differ in their conclusions about *when* common ground constrains referential domains. Some studies suggest that addressees are initially egocentric, considering potential referents of referring expressions, without regard for whether they are in common or privileged ground (Keysar, Lin & Barr, 2003; Keysar, Barr, Balin & Brauner, 2000). Other studies find that whereas addressees do attend to referents in privileged ground, they exhibit a strong and early preference for potential referents in common ground (Nadig & Sedivy, 2002; Hanna, Tanenhaus & Trueswell, 2003; Hanna & Tanenhaus, 2004).

The current studies take a different approach to exploring the role of common ground in real-time comprehension. Whereas previous studies have asked whether addressees exclude entities in privileged ground as potential referents, we examined whether addressees will shift their attention to entities in privileged ground when appropriate. One appropriate occasion is when an addressee is asked a question. An interlocutor who has been asked for information will typically provide it by introducing previously private or privileged information into the common ground; if it were already available to the questioner, after all, there would have been no need for the question. Thus, in a realistic dialogue where the participants expect each other to have both useful information to impart and a need for information that can be obtained from others, selective attention to privileged ground when being asked for information may reflect awareness of the knowledge states of others, rather than egocentricity.

To test this idea, we examined use of privileged ground information in an interactive conversation in which participants have a joint goal requiring them to exchange information. We hypothesized that the representations underlying use of common ground might be strongest in interactive conversation, especially when the participants have joint goals (Clark, 1992, 1996).

To create situations where a speaker is likely to refer to entities in an addressee's privileged ground, we created a 'targeted language game' (Brown-Schmidt, 2005, Brown-Schmidt & Tanenhaus, in press, Tanenhaus & Brown-Schmidt, in press) in which both participants have relevant private information, know that the other has relevant private information, and need to obtain information from the other to make progress in the game. We expected this setup to encourage the use of *wh*-questions (e.g. *What's in the upper right corner?*) to obtain the needed information; interpretation of such questions by the addressee could then be studied.

Experiment 1

Experiment 1 examined the role of shared and private information in the production of and interpretation of questions using a targeted language game designed to elicit specific linguistic constructions from conversational participants without explicitly controlling what they say. We hypothesized that speakers would use questions to ask about things in the addressee's privileged ground. If this hypothesis is correct, we can then examine the on-line interpretation of these questions. Upon hearing a question, we hypothesized that addressees would direct attention toward entities in the privileged ground, and away from entities in the common ground.

Method

Participants—We present data from twelve pairs of friends. Each participant was a native English speaker from Rochester, NY. Six additional participants were dropped from the analysis because they did not follow directions ($n=3$) or because of equipment problems ($n=3$). Participants were compensated \$7.50 per hour.

Materials and Procedure—Participants sat at a table facing each other. A game board was placed between the participants (see figure 1 and figure 2). One partner wore a head-mounted Applied Science Laboratories (ASL)-brand eye-tracker. The record of gaze, superimposed on a video-record of the scene from the eye-tracked participant's perspective, and both participants' voices, were recorded to a frame-accurate digital video recorder (Sony DSR-30) at 30 Hz.

The game board was a grid of 36 cubbyholes. Each cubbyhole contained a card with an identical animal picture on either side. There were three types of animals: pigs, horses and cows. Each animal had an accessory: shoes, hat or glasses. The 36 animals (4 of each type) were randomly assigned to a cubbyhole. The animals in 1/3 of the cubbyholes were visible to both participants, and thus were in common ground. The animals in 1/3 of the cubbyholes were only visible to the eye-tracked participant, and thus were in the eye-tracked participant's privileged ground. The remaining 1/3 of the cubbyholes were only visible to the non eye-tracked participant and thus were in the non eye-tracked participant's privileged ground.

The participants were instructed to re-arrange the animals so that adjacent cubbyholes had different types of animals and those animals had different accessories. Thus, two horses or two animals wearing glasses could not be next to each other. Participants were asked not to point, and were allowed to remove only two animals from the cubbyholes at a time. The game-board and game-pieces prevented participants from viewing the gaze or head position of their partner. The task lasted about 1h.

Game pieces were designed to elicit post-nominally modified referring expressions like *the pig with the hat*. The task was designed to elicit wh-questions like (1), which were the focus of our analyses.

(1) *What's below the cow with the hat?*

Results

Use of wh-questions—Each conversation was transcribed, and each wh-question that inquired about an animal on the game board was identified. Speakers primarily used wh-questions to inquire about animals they couldn't see (see table 1). The proportion of addressee privileged ground referents was significantly higher than either common ground referents, $t(11)=16.92$, $p<.0001$, Cohen's $d=6.89$, or speaker privileged referents, $t(11)=7.91$, $p<.0001$, $d=4.57$, whereas common and speaker privileged referents were equivalent, $t(11)=1.06$, $p=.31$.

Eye movement analyses—Addressees' eye movements during their partner's questions were analyzed in order to test the hypothesis that addressees will direct attention to privileged ground entities when interpreting questions.

We examined the type of cubbyhole (shared, addressee-private, speaker-private) that addressees looked at, relative to the onset of the critical noun (*cow*). At this point in the utterance, the addressee knows they are being asked about something next to a cow, but given the ambiguity in the game, there are multiple cows consistent with the question. If addressees distinguish shared from private information as they interpret a question, gaze should be preferentially directed toward their own (addressee-private) cubbyholes.

Figure 3 shows the proportion of fixations to shared, addressee-private and speaker-private cubbyholes as addressees interpreted wh-questions. Fixations were analyzed in a baseline region which captured fixations between 200ms before and 200ms after the onset of *cow*. The critical region was between +200 to +600ms after the onset of *cow* (signal driven fixations are not expected until approximately 200 ms after the onset of the critical word due to time needed

to program an eye movement). Figure 4 shows the proportion of fixations to shared, addressee-private and speaker-private cubbyholes across the two analysis regions.

An ANOVA with subjects as a random factor was used to analyze the proportion of fixations in the two regions; generalized eta-squared (Bakeman, 2005), provides an estimate of effect size. A main effect of cubbyhole type, $F(2,22)=10.18$, $p<.01$, $\eta^2_G=.45$, was qualified by a significant region x cubbyhole interaction, $F(2,22)=3.75$, $p<.05$, $\eta^2_G=.04$. Planned, one-tailed comparisons revealed that at baseline, the proportion of fixations to shared and addressee-private cubbyholes was equivalent, $p=.3$. Between the baseline and critical regions, fixations to shared cubbyholes dropped and fixations to private cubbyholes increased (p 's<.05, d 's>.37).

Conclusion

The results clearly demonstrate that speakers primarily use wh-questions to ask for information about entities in an addressee's privileged ground. The results also provide suggestive evidence that addressees in an unscripted conversation can use information about sentence type, in combination with knowledge about what is and is not in the common ground, to quickly guide attention towards private information when interpreting questions. Experiment 2 was designed to examine the interpretation of wh-questions using a simpler display in a more controlled setting while preserving the interactive nature of the task.

Experiment 2

We modified the game so that each participant had a conversation with the experimenter. Periodically, the experimenter used scripted utterances, which provided the experimental conditions of interest. We again tested the hypothesis that addressees would interpret questions as asking about something in their privileged ground. We also compared the relative strength of two ways that entities can be entered into common ground: linguistic and visual co-presence. We hypothesized that in a conversation, linguistic co-presence might play a stronger role than visual co-presence because discourse referents are often not visually co-present in everyday conversation.

Method

Participants—48 native English speakers from Urbana-Champaign, IL participated in this experiment. Data from an additional 5 participants were excluded due to equipment problems. Participants were compensated either \$8 or course credit.

Materials and Procedure—The participant and experimenter were seated at separate computers, facing away from each other. The participant wore a head-mounted Eye-Link II eye-tracker. The participant's gaze was recorded at 250hz and saved to disk along with an audio record of both voices.

Across 36 trials, the experimenter and participant viewed 3×3 grids with animals on their respective computer screens (see figure 5–figure 6). The animals were identical to those used in Experiment 1. The animals in 1/3 of the squares on the grid were visible on both the experimenter's screen and the participant's screen, thus these animals were in common ground. The common ground squares were distinguished by a white background. The animals in 1/3 of the cubbyholes that were only visible to the participant, and thus in the participant's privileged ground, had a gray background. The remaining 1/3 of the cubbyholes were only visible to the experimenter, and thus in the experimenter's privileged ground.

On each trial, the task was to check whether adjacent squares had different types of animals with different accessories. As in Experiment 1, two horses could not be next to each other, and

two animals wearing glasses could not be next to each other. If the experimenter and participant identified an error, they moved on to the next trial. The experiment typically lasted 45min.

On 24 of the 36 trials, the experimenter used a scripted pair of utterances, illustrated in examples 2–5. These critical trials never contained errors. The twelve filler trials were unscripted; eight contained an error.

We used a point-of-disambiguation manipulation (Eberhard, Spivey-Knowlton, Sedivy & Tanenhaus, 1995). Half of the critical trials had a late point-of-disambiguation. On these trials the experimenter either made a statement or asked a question. Then the experimenter asked the critical question which always had the form of “*What’s (above/below) the x with the y?*”. Crucially, at the first noun (*cow* in examples 2a–b), the critical question was temporarily consistent with both the target of the question (e.g. the pig with glasses in figure 5) and the competitor (e.g. the horse with shoes). The point-of-disambiguation came at the second noun (*hat*).

The remaining trials had a potentially earlier point-of-disambiguation because the competitor was in the either the linguistic or visual common ground. For early-linguistic trials (n=6), the experimenter asked the participant what was in the competitor square immediately before the critical question (figure 5/example 3). Thus, the participant’s response to this set-up question brought the competitor into the linguistic common ground. For early-visual trials (n=6), we brought the competitor into the common ground using visual co-presence (figure 6/example 4).

2a. Late disambiguation trial, example 1 (see figure 5)

There’s a pig with shoes in the middle. (e.g. from the experimenter’s view)

What’s above the cow with the hat?

2b. Late disambiguation trial, example 2 (see figure 5)

What’s in the bottom middle?

(participant: *A horse with glasses.*)

What’s above the cow with the hat?

3. Early disambiguation trial, linguistic condition (see figure 5)

What’s in your top left corner?

(participant: *A horse with shoes.*)

What’s above the cow with the hat?

4. Early disambiguation trial, visual condition (see figure 6)

What’s in the bottom middle?

(participant: *A horse with glasses.*)

What’s above the cow with the hat?

Each target picture was rotated through the experimental conditions across four lists. Trials were presented in one of two random orders. A prosodic analysis on a subset of trials (25% of all critical trials) determined there were no pitch, duration or intensity differences between conditions for any of the words preceding the point-of-disambiguation. The critical noun phrase was never globally ambiguous. Thus, in figure 5–figure 6, there was only a single cow with a hat in the scene.

Results and Discussion

Eye movement analyses—We compared the eye movements that addressees made as they interpreted wh-questions with an early vs. late point-of-disambiguation. If addressees use the distinction between shared and private information when interpreting wh-questions, addressees should initially make fewer competitor fixations and more target fixations in the early compared to the late point-of-disambiguation conditions.

Eye movements were analyzed in terms of target advantage (Arnold, Eisenband, Brown-Schmidt & Trueswell, 2000), which is calculated as the proportion of fixations to the target–competitor (see figure 7). In figure 5 and figure 6, the target was the pig with glasses, and the competitor was the horse with shoes.

The target advantage scores for the early and late point-of-disambiguation conditions were analyzed across three 400ms time regions in a condition x region ANOVA. The first region (baseline), captured fixations between –200ms and +200ms following critical noun onset (*cow*). The second and third regions captured fixations from +200ms to +600ms, and +600ms to +1000ms after critical noun onset, respectively. Figure 8 shows the average target advantage scores for the early and late point-of-disambiguation conditions across the three regions. The main effect of condition was due to significantly larger target advantage scores for trials with an early point-of-disambiguation, $F(1,47)=5.35, p<.05, \eta^2_G=.01, F(2,18)=10.76, p<.05, \eta^2_G=.11$, demonstrating that addressees used common ground information to identify the referential domain of the question. An epoch by condition interaction was significant in the participants analysis only (a small number of items limited power in the items analysis), $F(2,94)=3.17, p<.05, \eta^2_G=.01, F(2,16)=1.74, p=.21, \eta^2_G=.08$, suggesting that the condition effect emerged over time. Planned, one-tailed comparisons demonstrated that at the baseline region, early and late conditions did not differ, $t's<.26, p's>.8$. At the second and third regions, the early point-of-disambiguation condition had significantly higher target advantage scores than the late condition, region 2: $t(47)=3.08, p<.01, d=.39, t(8)=3.97, p<.01, d=1.26$, region 3: $t(47)=2.71, p<.01, d=.23, t(8)=1.83, p<.05, d=1.21$. A direct comparison of the early linguistic and visual conditions revealed no significant differences in target advantage at any region, $t's<1.2, p's>.28$.

The results clearly demonstrate that addressees can use the distinction between shared and private information as they interpret language on-line. When common ground information restricted the domain of interpretation, addressees directed attention toward the target and away from the competitor referent within 400ms of critical noun onset. When common ground information did not constrain the domain of interpretation, fixations to the target and competitor did not diverge until 700ms, which corresponds to hearing the disambiguating noun (*hat*, which occurred about 450 ms after critical word onset). We found no evidence that the timing of addressee's use of information about common ground differed as a result of whether it was established by visual or linguistic co-presence. The results for linguistic-co-presence are to the best of our knowledge the first demonstration that information established by linguistic co-presence can influence whether an addressee considers a potential referent to be common or privileged.

We note that (in both experiments) the preference for the privileged ground target emerged at the noun, rather than at *what* or *what's above*. This is probably related to the variety of uses of the word *what* (e.g. *What's above it is a cow, which won't work.*), in combination with the complexity of the scene. In addition, in previous studies which demonstrated immediate effects at a preposition (e.g., Chambers, Tanenhaus, Eberhard, Filip & Carlson, 2002, Eberhard et al., 1995), the preposition introduced a goal for an action involving a theme or two potential themes that been established *prior* to the preposition (e.g., *the five of hearts that is above...*). In contrast, the preposition in the current study preceded a goal that was being introduced as a reference

point (anchor) for a questioned-theme. Whether this distinction is important is an issue for future research.

General Discussion

In everyday conversation, speakers routinely both provide information to their interlocutors and seek information from them. Recognizing when each is called for seems to require, in addition to linguistic knowledge, inferences about who is likely to know what. We examined the on-line interpretation of questions in interactive conversations based on the assumption that interlocutors are most likely to make use of observations about each other's shared (and different) task-relevant knowledge during collaborative tasks that establish common ground through interaction. Unlike Keysar et al. (2003) and Keysar et al. (2000), we did not find evidence for an early stage of language processing in which addressees ignore the distinction between shared and private information. Rather, our results are consistent with studies that find immediate effects of common ground (Hanna et al., 2003; Hanna & Tanenhaus, 2004; Nadig & Sedivy, 2002). However, unlike those studies, which used imperatives (directing the addressee's attention to referents in common ground), we used information questions, which focused the addressee's attention on referents in private ground. Our results demonstrate that how the language processing system uses representations of shared and private knowledge is likely to be strongly influenced by the form of the uttered sentence, with direction and strength of the effects determined by the predisposition of the sentence type uttered together with factors specific to the discourse context.

Recent work on the use of perspective in on-line language processing has made the simplifying assumption that an egocentric-first system will initially ignore differences between shared and privileged ground, whereas a system sensitive to these differences will restrict itself to information that is in common ground, thereby filtering out privileged knowledge. However, our results demonstrate that this assumption is clearly incorrect. Different utterance forms differ systematically in their capacity for introducing referents likely to be shared, private, or of uncertain provenance. Thus, an addressee who is sensitive to the perspective of the speaker will attend to both shared and private information, at different times and in different linguistic contexts.

What remains unclear is why some studies find that addressees initially ignore utterance-relevant differences between privileged and common ground. Although a detailed discussion of this question goes beyond the current work we have two hypotheses. First, egocentric behavior is most likely to emerge when the critical word matches a referent in the addressee's privileged ground better than any common ground referent (for discussion see Hanna et al., 2003; Heller, Grodner & Tanenhaus, 2007). Second, we believe that effects of perspective are likely to be strongest in tasks where participants have joint goals, common ground is established collaboratively, and exchange of information is negotiated by both parties.

Acknowledgment

This research was partially supported by NIH grant HD 27206 to M. K. Tanenhaus. Thank you to Courtney Pooler and Carol Faden for their help with data analysis.

References

- Arnold JE, Eisenband JG, Brown-Schmidt S, Trueswell JC. The immediate use of gender information: eyetracking evidence of the time-course of pronoun resolution. *Cognition* 2000;76:B13–B26. [PubMed: 10822045]
- Bakeman R. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods* 2005;37:379–384. [PubMed: 16405133]

- Brown-Schmidt, S. Dissertation Abstracts International: Section B: The Sciences and Engineering. 66. 2005. Language processing in conversation (Doctoral dissertation, University of Rochester, 2005); p. 5079
- Brown-Schmidt S, Tanenhaus MK. Real-time investigation of referential domains in unscripted conversation: a targeted language game approach. *Cognitive Science*. in press
- Chambers CG, Tanenhaus MK, Eberhard KM, Filip H, Carlson GN. Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language* 2002;47:30–49.
- Clark, HH. *Arenas of language use*. Chicago: University of Chicago Press; 1992.
- Clark, HH. *Using language*. Cambridge, UK: Cambridge University Press; 1996.
- Eberhard KM, Spivey-Knowlton MJ, Sedivy JC, Tanenhaus MK. Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 1995;24:409–436. [PubMed: 8531168]
- Hanna JE, Tanenhaus MK. Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 2004;28:105–115.
- 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. University of Rochester; 2007. Manuscript submitted for publication
- Keysar B, Lin S, Barr DJ. Limits on theory of mind use in adults. *Cognition* 2003;89:25–41. [PubMed: 12893123]
- Keysar B, Barr DJ, Balin JA, Brauner JS. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Sciences* 2000;11:32–38.
- 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, RC. Assertion. In: Cole, P., editor. *Syntax and semantics: Pragmatics*. Vol. 9. New York, NY: Academic Press; 1978. p. 315–332.
- Tanenhaus MK, Brown-Schmidt S. Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*. in press
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. *Science* 1995;268:1632–1634. [PubMed: 7777863]

Figure 1.
An image of the game board from the eye-tracked participant's perspective. The crosshair indicates where the eye-tracked participant is gazing.

Figure 2.
A schematic of the experimental setup.

Figure 3. Proportion of fixations to shared, speaker-private and addressee-private entities over time. Data are centered at 0ms= critical noun onset (e.g. *cow* in: *What's below the cow with the hat?*).

Figure 4. Proportion of fixations to shared, speaker-private and addressee-private entities, averaged across two 400ms time regions. Error bars indicate standard error of by-participant means.

Figure 5. Example display for the late point-of-disambiguation trials, and the early point-of-disambiguation, linguistic condition. Display is from the eye-tracked participant's perspective.

Figure 6. Example display for early point-of-disambiguation trials, visual condition. Display is from the eye-tracked participant's perspective. Note that the key difference between Figure 5 and Figure 6 is that the competitor (the horse with shoes) is in the visual common ground in Figure 6.

Figure 7.

Target advantage scores (proportion of target fixations-proportion of competitor fixations) for the early and late point-of-disambiguation conditions, centered at the onset of the critical noun (e.g. *cow* in: *What's above the cow with the hat?*).

Figure 8. Target advantage scores for early and late point-of-disambiguation conditions, averaged across three 400ms regions. Error bars indicate standard error of by-participant means.

Table 1

The proportion of wh-questions (out of 166 total utterances) that inquired about common, speaker-privileged, and addressee-privileged entities. The 95% confidence interval of the by-participants mean is shown in parentheses.

Common Ground	Speaker's Privileged Ground	Addressee's Privileged Ground
.00 (.01)	.06 (.12)	.93 (.12)