Cognition 108 (2008) 831–836

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Brief article

The role of perspective in identifying domains of reference ☆

Daphna Heller [a,*], Daniel Grodner [a,b], Michael K. Tanenhaus [a]

[a] Department of Brain & Cognitive Sciences, University of Rochester, Meliora Hall, Box 270268, Rochester, NY 14627, USA
[b] Department of Psychology, 500 College Avenue, Swarthmore, PA 19081, USA

ARTICLE INFO

Article history:
Received 24 July 2007
Revised 26 April 2008
Accepted 30 April 2008

Keywords:
Common ground
Reference resolution
Perspective
Eye-tracking
Language

ABSTRACT

We used the contrastive expectation associated with scalar adjectives to examine whether listeners are sensitive to the distinction between common and privileged information during real-time reference resolution. Our results show that listeners used this distinction to narrow the set of potential referents to objects with contrasts in common ground from the earliest moments. These results extend previous evidence that ground information influences real-time language processing by showing that the distinction between common and privileged information is used without being triggered by unusual circumstances.

1. Introduction

Formal theories of conversation assume that interlocutors are sensitive to each other's knowledge and how it differs from their own. Accounts of the felicity conditions for making assertions, asking questions, and using referring expressions often appeal to the distinction between information in the interlocutors' common ground and information that is privileged to the speaker or the addressee. For example, imperatives typically refer to information in common ground, whereas questions inquire about information that is privileged to the addressee.

Determining what is common and what is privileged requires computing information from multiple sources, including the physical and the linguistic context. Therefore, these computations may be too slow or burdensome for real-time processing. Support for this view comes from Keysar, Barr, Balin, and Brauner (2000), who examined the time-course of perspective-taking using visual-world eye-tracking (Cooper 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). A confederate speaker instructed participants to manipulate objects in cubbyholes. Some objects were visible to both interlocutors and were thus in common ground by physical co-presence (Clark & Marshall, 1981). Others were visible only to listeners, and were thus in their privileged ground. Participants followed instructions like "pick up the small candle" where the display contained two shared candles that differed in size and a third smallest candle which was privileged to the listener. Listeners were more likely to first look at the privileged candle and sometimes even reached for it, before identifying the intended referent. Keysar et al. concluded that listeners' reference resolution proceeds initially relative to their egocentric perspective, ignoring the distinction between common and privileged ground.

Other studies have found early effects of ground. Nadig and Sedivy (2002) and Hanna, Tanenhaus, and Trueswell (2003) compared conditions in which a referring expression was ambiguous between two objects in common ground with conditions in which one of these objects was privileged. In Hanna et al. Experiment 1, for example, the confederate instructed listeners to "put the blue circle

☆ We are grateful to Jenni Fasching, Shira Schwartz, Ali Horowitz, and especially Dana Subik and Rebekka Puderbaugh for their help in creating the stimuli, collecting and coding the data. We thank the editor Vic Ferreira, and the reviewers for insightful comments that much improved this article. This research was partially supported by NIH Grant HD-22067 to M.K. Tanenhaus.

* Corresponding author. Tel.: +1 585 273 1191; fax: +1 585 273 1088.
E-mail address: dheller@ling.rochester.edu (D. Heller).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.04.008

above the red triangle", comparing conditions with two shared red triangles to conditions where one of them was privileged to the listener. When both objects were in common ground, listeners were equally likely to look at either, but when one object was privileged, listeners were more likely to look at the shared object from the earliest moments and were faster to choose it (although they were more likely to look at a privileged competitor than at an unrelated privileged object). In these studies, the referring expressions were globally ambiguous and thus infelicitous from the listener's perspective (also see Hanna & Tanenhaus, 2004). Since the ambiguity can only be resolved by appealing to ground information, these findings are consistent with an "egocentric-first" heuristic, where ground information is used only when triggered by unusual circumstances, such as the infelicity caused by global ambiguity (Keysar, Lin, & Barr, 2003).

The current study asks whether listeners use ground information when there is nothing unusual in the instructions that might trigger them to rely on this kind of information. Participants played the role of addressee in a referential communication task while their eye movements were monitored. Common ground was established by physical co-presence. We exploited the contrastive function associated with scalar adjectives (Sedivy, 2003), employing it in a point-of-disambiguation manipulation (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995). For example, in "pick up the big duck", the scalar adjective "big" creates the expectation that the speaker will refer to the big member of a pair contrasting in size, rather than an object which is big in an absolute sense. When the visual context contains a size contrast, participants will often fixate on the big member of the contrast even before encountering information from the noun (Sedivy, Tanenhaus, Chambers, & Carlson, 1999). This allowed us to use instructions that were unambiguous, thereby avoiding any infelicity or other unusual circumstances that might encourage listeners to strategically use ground information.[1]

We compared displays with one size contrast, which have an early point-of-disambiguation with displays containing two size contrasts, where disambiguation is not expected until the noun. We also manipulated whether one object was in the listener's privileged ground. In displays with two size contrasts, this object was the competitor-contrast. The full design is depicted in Fig. 1.

If listeners process egocentrically, the target in both conditions with two contrasts should not be identified until the noun is encountered, independent of the ground status of the competitor-contrast. If, however, listeners

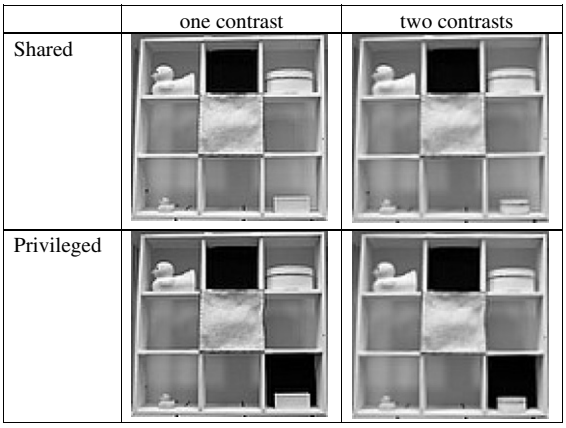

Fig. 1. Example displays for the instruction "*pick up the big duck*". In the one contrast conditions the competitor-contrast (small box) was replaced by an unrelated object (a bar of soap). In the privileged conditions, these objects were only visible to the listener (squares backed by a black cloth were only visible to the listener).

encode whether information is common or privileged and use this distinction in real-time, the adjective should allow listeners to anticipate the target when the competitor-contrast is in their privileged ground, because they are not expecting the speaker to use a scalar adjective in referring to the competitor for which the speaker has no contrast.

Previous discussions have often contrasted an egocentric-first heuristic with a common-ground heuristic, where listeners interpret referring expressions relative to common ground, ignoring information in their privileged ground. However, as pointed out earlier, some types of utterances typically refer to information in common ground whereas others typically refer to privileged information (Brown-Schmidt, Gunlogson, & Tanenhaus, 2008). Therefore, optimal listeners should be sensitive to what information is shared and what information is privileged to them, as well as what information might be privileged to the speaker. This contrasts with the egocentric-first heuristic, where listeners initially ignore perspective information altogether, and with the common-ground heuristic where listeners focus solely on mutual information.

We use looks to privileged objects to assess whether listeners are ignoring information in privileged ground, as suggested by the common-ground heuristic. In particular, when a referent has a contrast, listeners will typically look at its contrasting object after identifying it (Sedivy et al., 1999). If listeners are aware of the contents of the information in privileged ground, we expect them to look at the privileged object more when it provides a potential contrast to another object in the display than when it is unrelated.

2. Methods

2.1. Participants

We present data from 16 participants, all native speakers of English from Rochester, NY. Four additional participants were excluded from analysis because of equipment problems or mistakes in the procedure. Participants were paid $15.

[1] Preliminary evidence that ground information is used in the absence of global ambiguity comes from Hanna et al. (2003) Experiment 2, which exploited the contrastive function associated with adjectives like *empty*. In this experiment, the objects were visible to the listener only, and were wrongly described by the experimenter to the confederate speaker in a way that created mismatching perspectives. The results showed that listeners adopted the speaker's perspective when interpreting the speaker's instruction. It is possible, however, that the experimenter's unusual error encouraged listeners to strategically adopt the speaker's perceptive. Moreover, it is possible that listeners adopted the speaker's perspective because the speaker's perspective was incompatible with their own perspective – this contrasts with all other studies discussed in this paper.

2.2. Materials

Critical instructions were "pick up the [scalar adjective] [noun]". Two factors were manipulated in constructing displays: number of contrasts (one vs. two) and ground (shared vs. privileged). Displays in the two contrasts conditions contained two pairs of size contrasting objects, e.g. a big duck (target) and a small duck (target-contrast), a big box (competitor) and a small box (competitor-contrast). Care was taken to match the pairs for size and visual salience. One contrast displays contained an unrelated object (distracter) instead of the competitor-contrast, which was matched to the competitor-contrast in size and visual salience. Item sets are presented in Appendix A. The ground factor manipulated whether all four objects were visible to both interlocutors or whether there was one object that was visible only to the listener.

Sixteen experimental sets were constructed for each of the four conditions. One condition was assigned to each of four lists and rotated across participants using a modified Latin square design. Each participant saw one version of each item set.

Fillers were constructed to avoid contingencies that might bias listeners towards the target. Twelve fillers had displays similar to experimental items, but the target was not a member of a shared contrast. In four of these, a scalar adjective was used to refer to a singleton object. In another four, a scalar was used for an object whose contrast was privileged to the listener. These fillers eliminated contingencies between the use of scalar adjectives and the presence of a (shared) contrast. Four additional displays in which the referent was a singleton also contained a size contrast with a privileged member. Eight displays had one color contrast: a color adjective was used to refer to either a member of this contrast or to a singleton. Eight displays had no contrasts and used no adjectives. The resulting 48 trials were presented in two orders, yielding eight lists.

2.3. Procedure

Participants were told that the purpose of the experiment was to investigate how people cooperate on a collaborative task when their perspectives differ. They were truthfully informed that the (female) speaker was a lab assistant who was naive to the goals of the experiment. Participants did not know that the instructions were partially scripted.

Participants sat at a table facing the speaker with a 3×3 vertical display between them. The upper two cubbyholes in the center were covered, so interlocutors could not see each other's face. Participants' eye movements were tracked using a head-mounted ASL 5000 eye-tracker. The gaze of the participant, superimposed on a video record of the scene, and both voices were recorded to a Sony DSR-30 digital video recorder at 30 Hz.

At the start of each trial, the speaker covered the four corner cubbyholes from her side of the display, so their contents were only visible to the participant. The participant was handed a bag containing four objects and a photograph directing him where to place the objects. The speaker faced the wall until the participant informed her that the objects were in place. Then, she was handed a photograph showing the final state of the display from her perspective, indicating which covers to remove and the final location of the target.

The first part of the instruction was "pick up the [referring expression]". Unbeknownst to the participant, the speaker's card provided the referring expression. She then improvised the moving instruction, e.g. "... and move it one space down". To control the referring expressions used throughout the experiment, the speaker was instructed not to refer to other objects, except in three fillers. In these fillers, the display contained a size contrast with one privileged member, and the speaker referred to the shared member using a bare definite, e.g. "... and put it under the bottle". The goal was to draw the participant's attention to the fact that the speaker could not see the privileged objects. One such filler was presented during practice, and the other two during the first two-thirds of the experiment.

Six of the 48 trials were used for practice. Three were used to familiarize the participant with the task; the other three provided the participant an opportunity to play the role of speaker.

3. Results

Debriefing questionnaires asked participants to describe what the experiment was investigating and whether they noticed anything odd. None of the participants suspected that the instructions were scripted. Several participants suspected that the speaker knew what the privileged objects were on *just* those fillers where the scalar was infelicitous because the target-contrast was privileged.[2]

Eye movements were analyzed from the video records using a VCR with synchronized audio and video channels. Fixations were coded for which cubbyhole participants were looking at beginning 200 ms before the onset of the adjective and ending when the participant reached for the target. One trial was excluded because the participant reached for the wrong object.

Fig. 2 plots the proportion of fixations to the target over time. To focus on speech-driven fixations, we excluded trials where participants were already looking at the target at the beginning of the trial. Looks to the target in the two-shared condition began to lag behind looks to the target in the other three conditions 200–300 ms after adjective onset, indicating that in the two-shared condition disambiguation was delayed until the noun. Crucially, looks to the target were not delayed in the two-privileged condition, as would be predicted if listeners were interpreting the instruction egocentrically.

We calculated the ratio of proportion of fixations to the target over the sum of proportion of fixations to the target and its competitor. To avoid problems inherent to proportional data, participant and item averages were quasi-logit transformed (Agresti, 2002; Jaeger, in

[2] If participants has suspected that the speaker usually knew what the privileged objects were, then contrary to our findings, there would not have been an effect of ground.

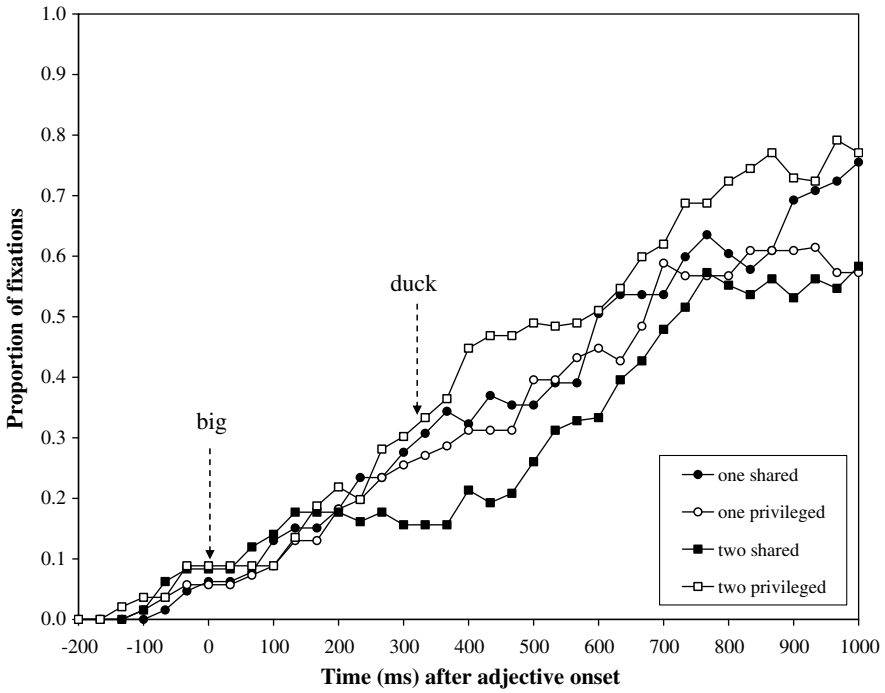

Fig. 2. Proportion of fixations to the target in the four conditions. Trials are aligned to the onset of the scalar adjective, e.g. "big" at 0 ms. The average onset of the noun, e.g. "duck", is also marked on the graph (320 ms).

press) prior to ANOVA analysis.[3] Fig. 3 plots target advantage ratios over time.[4]

We compared target advantage ratios for fixations in the baseline region, between 200 ms before and 200 ms after adjective onset, and in the adjective region, spanning 200 ms after adjective onset to 200 ms after the average noun onset (a window of 333 ms). There were no differences among the conditions (Fs < 1) in the baseline region. This was expected given estimates of 150–200 ms to program and launch a saccade (Matin, Shao, & Boff, 1993). In the adjective region, there was a reliable interaction between number of contrasts and type of ground ($F1(1,15) = 4.77$, MSE = 2.97, $p < .05$; $F2(1,15) = 5.92$, MSE = 1.44, $p < .05$). Planned comparisons established that the target advantage ratio was reliably higher for the one-shared condition compared to the two-shared condition ($F1(1,15) = 3.82$, MSE = 2.01, $p < .05$; $F2(1,15) = 3.31$, MSE = 1.75, $p < .05$), confirming that multiple contrasts led to later identification of the target. Crucially, the target advantage ratio was reliably higher for the two-privileged condition compared to the two-shared condition ($F1(1,15) = 12.8$, MSE = 1.69, $p < .01$; $F2(1,15) = 7.4$, MSE = 2.62, $p < .01$).[5] These results clearly demonstrate that, contrary to the predictions of the egocentric-first heuristic, the common vs. privileged status of objects influenced listeners' expectations about the speaker's referring expressions. We note that there is a numerical, but not statistical, difference in target advantage ratios in the two-privileged condition, compared to conditions with one contrast.[6]

We now turn to the question of whether listeners ignored information in privileged ground, as suggested by the common-ground heuristic. Recall that after identifying a member of a size contrast as the referent, listeners will often look at the contrasting object (Grodner & Sedivy, in press; Sedivy et al., 1999). If participants were aware that the privileged object was a potential contrast, we would expect more looks to this contrasting object compared with an unrelated distracter. We tested this prediction by comparing looks to the privileged objects in the two-privileged

[3] See Jaeger, in press for discussion of the advantages of quasi-logit compared to an arcsine transform.

[4] Early fixation proportions are lower than .5 because trials where the participant was already fixating on the target at the beginning of the trial were excluded.

[5] A reviewer suggested looking at new fixations instead of proportions. For the three early point-of-disambiguation conditions (one-shared, one-privileged and two-privileged), 67% of the new fixations launched during the adjective region were to the target. In the late point-of-disambiguation condition (two-shared), fixations to the target *and* its same-size competitor constituted 69% of the new fixations.

[6] It was suggested during the reviews that whenever the privileged object contrasted with a shared object, participants might have followed a strategy of ignoring this size contrast altogether, focusing their attention on the shared size contrast as they waited for the instruction. This might have accounted for the numerical (but not statistical) difference in target looks in the two-privileged condition, compared to conditions with one contrast. To evaluate this option, we calculated the ratio of looks to the shared size contrast (target + target-contrast) over all three shared objects (target + target-contrast + competitor) in the baseline region, i.e. before the instruction. If participants were strategically focusing on the shared size contrast in the two-privileged condition, we would expect a ratio that is higher than chance (.67). However, the ratios were similar: (.63) for the two-shared condition and (.66) for the two-privileged condition. Furthermore, the target advantage ratios in this region (target over [target + competitor]) were also similar for the two-shared condition (.47) and the two-privileged condition (.44), indicating that the measure in our main analyses was appropriate.

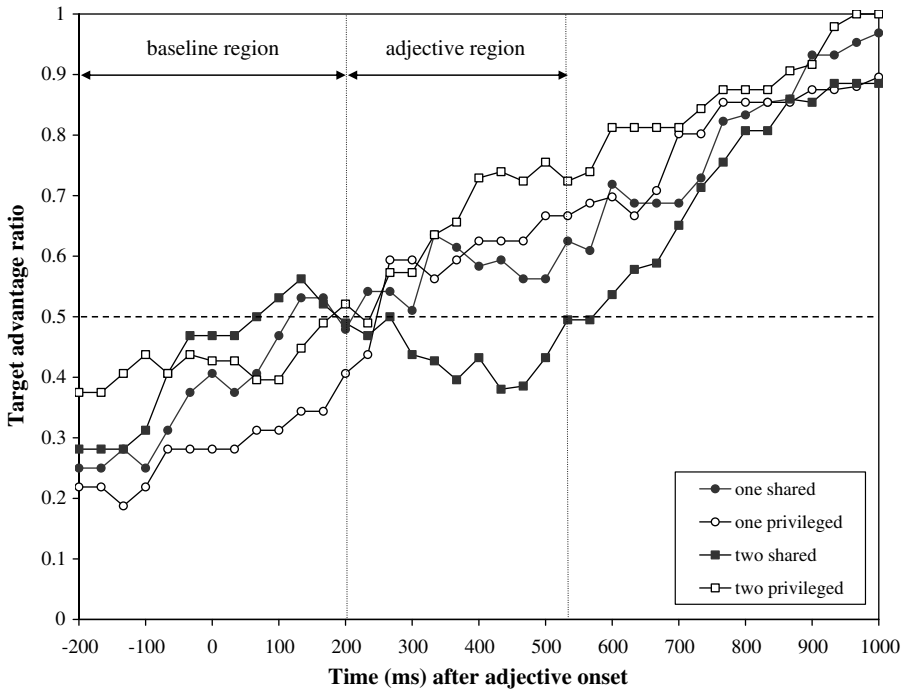

Fig. 3. Target advantage ratios in the four conditions: fixations to the target over the sum of fixations to the target and its same-size competitor. The onset of the adjective is at 0 ms and the average onset of the noun is at 320 ms. Average regions of analysis are marked on the graph.

and the one-privileged conditions. As expected, looks did not differ across the two conditions during the baseline region or the adjective region. We expected to find the effect during the noun region, spanning 200 ms after the average noun onset to 200 ms after the average noun offset (a 533 ms window), since this is just after the referent is identified and where such effects have been observed in previous work. The proportion of fixations to the privileged object was significantly higher in the two-privileged condition than in the one-privileged condition ($F1(1,15) = 4.02$, MSE = 2.01, $p < .05$; $F2(1,15) = 3.76$, MSE = 2.33, $p < .05$), confirming that listeners looked more at the privileged object when it provided a potential contrast than when it was unrelated. This indicates that listeners were aware of the identity of the privileged object and did not block privileged ground information from attention (see Wardlow Lane, Groisman, & Ferreira, 2006, who present evidence for *speakers* not blocking privileged information).

4. Conclusions

We observed an early point of disambiguation for conditions with one size contrast compared to the condition with two shared size contrasts, replicating previous results where scalar adjectives create an expectation that the speaker will refer to an object with a size contrast (Sedivy, 2003; Sedivy et al., 1999). Reference resolution was also early when an object privileged to the listener created a second contrast, indicating that listeners' expectations about the speaker's referring expressions were based on shared contrast(s) alone. Our findings are consistent with Nadig and Sedivy (2002), Hanna et al. (2003), and Hanna and Tanenhaus (2004) who observed an early effect of ground (see also Wu & Keysar, 2007). Crucially, however, our results demonstrate that listeners use the distinction between common and privileged ground when processing felicitous referring expressions which are not globally ambiguous. Since most referring expressions are temporarily ambiguous as the utterance unfolds, temporary ambiguity is not expected to trigger special use of ground information. Therefore, these results cannot be explained by an egocentric-first heuristic.

We provided evidence that although listeners may restrict their referential domain to information in common ground when appropriate, they are nonetheless aware of information in privileged ground. We thus claim that taking perspective does not mean adopting a common-ground heuristic, whereby attention is only given to mutual information, but rather being aware of the common or privileged status of information. We propose that listeners encode the status of information as common or privileged and use this distinction in real-time reference resolution.

How can we reconcile the results of studies that find early effects of ground, with those that do not? Goodness of fit to the speaker's referring expression is typically one of the most reliable of the probabilistic cues available to the listener for identifying the intended referent. Recall that in Keysar et al. (2000) study, listeners initially selected a privileged object to be the target, seeming to ignore the distinction between common and privileged ground. Importantly, in that study the privileged object was always a better fit to the referring expression than the intended referent in common ground (e.g. the smallest candle vs. the

medium candle for "the small candle"). If listeners encode information in privileged ground, a privileged object that best matches the referring expression is likely to attract their attention. In the current study, by contrast, goodness of fit to the referring expression and ground information both pointed to the same referent, allowing us to observe the real-time integration of ground information.

More generally, we suggest that apparent egocentric behavior is most likely to be observed when certainty about the status of information as common or privileged is low, in which case ground information will not be perceived as a reliable cue, or when another strong constraint, such as goodness of fit to the speaker's referring expression, conflicts with ground. In future research, it will be important to evaluate this claim by quantifying and manipulating these factors – a strategy that has proved fruitful in evaluating constraint-based approaches in other domains.

Appendix A. List of items

Target	Target-contrast	Competitor	Competitor-contrast	Distracter
Big hairclip	Small hairclip	Big scissors	Small scissors	Eraser
Big tape dispenser	Small tape dispenser	Big stamp	Small stamp	Nail polish
Big gluebottle	Small gluebottle	Big can	Small can	Mug
Big funnel	Small funnel	Big tupperware	Small tupperware	Salt shaker
Big duck	Small duck	Big box	Small box	Bar of soap
Big stapler	Small stapler	Big car	Small car	Pinecone
Big screwdriver	Small screwdriver	Big block	Small block	Frog
Big bow	Small bow	Big candle	Small candle	Pear
Small bowl	Big bowl	Small pipe	Big pipe	Tongs
Small pot	Big pot	Small 8 ball	Big 8 ball	Jar
Small cup	Big cup	Small deodorant	Big deodorant	File card box
Small scoop	Big scoop	Small lego	Big lego	Egg
Small spring	Big spring	Small sharpie	Big sharpie	Comb
Small lock	Big lock	Small bird's nest	Big bird's nest	Mouse
Small basket	Big basket	Small sponge	Big sponge	String
Small balloon	Big balloon	Small post-its	Big post-its	Can opener

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New Jersey: John Wiley and Sons Inc.

Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition, 107*, 1122–1134.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. H. Joshi, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84–107.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye-movements as a window into spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409–436.

Grodner, D., & Sedivy, J. (in press). The effects of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*, 43–61.

Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science, 28*, 105–115.

Jaeger, T. F. (in press). Better categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32–37.

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

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics, 53*, 372–380.

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

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental processing through contextual representation: Evidence from the processing of adjectives. *Cognition, 71*, 109–147.

Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research, 32*, 3–23.

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

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don't talk about pink elephants! Speakers' control over leaking private information during language production. *Psychological Science, 17*(4), 273–277.

Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science, 18*, 600–606.