# The perceptual foundation of linguistic context

**Francis Mollica (fmollica@bcs.rochester.edu)**
Department of Brain and Cognitive Sciences, Meliora Hall, Box
Rochester, NY 14627 USA

**Steven T. Piantadosi (spiantadosi@bcs.rochester.edu)**
Department of Brain and Cognitive Sciences, Meliora Hall, Box
Rochester, NY 14627 USA

**Michael K. Tanenhaus (mtan@bcs.rochester.edu)**
Department of Brain and Cognitive Sciences, Meliora Hall, Box
Rochester, NY 14627 USA

**Abstract**

Context plays a ubiquitous role in language processing. For the most part, work in language processing investigates the effects of context without investigating questions about what determines a context. For example, interpretation of any referential expression must take into account the notion of a referential domain. Here we investigate the influence of perceptual cues in establishing a referential domain, or linguistic context. We demonstrate that people use perceptual cues to establish a linguistic context; the influence of perceptual cues is gradient with respect to cue magnitude; and the contribution of a perceptual cue in constructing a linguistic context is not an effect of attention or salience. We provide these results as a first step toward developing a formal model for the establishment of linguistic context.

**Keywords:** language processing; reference resolution; linguistic context

## Introduction

Context plays a ubiquitous role in language processing. One relatively well understood case is the role that a referential domain plays in generating and understanding definite referring expressions. Definite referring expressions are commonly understood as picking out a uniquely identifiable referent. As an example, consider the interpretation of a request such as, *"Can you pass me the red wine?"* uttered at a small table in a crowded NYC restaurant. If there were carafes of the house wines at the table, one red and one white, then *"the red wine"* would clearly be intended, and understood, to refer to the carafe of red at the table. However, in a NYC restaurant, there would likely be numerous carafes of the house red at the table; there might even be a carafe of red wine that is at an adjacent table closer to one of your dinner companions than your own carafe.

Why are these carafes not potential referents, and why is the use of a definite article felicitous when there is clearly more than one carafe of red? The answer is that any referring expression must be interpretable with respect to a relevant referential domain, the context that defines the set of available referents. There is a body of research that seeks to establish what constrains referential domains, investigating the role of general world knowledge (e.g., our table versus someone elses table), goal-specific constraints—including action-based affordances (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers, Tanenhaus, & Magnuson, 2004), and the common ground that is shared between interlocutors because of community membership, physical co-presence and linguistic co-presence—i.e., the history of the discourse (Clark & Marshall, 1978).

Some aspects of what constitutes a likely referential domain might, however, be based upon much simpler, basic perceptual foundations. These basic perceptual effects might serve as the substrate upon which more complex contexts are built. A hint that this might be the case comes from a study by Klein, Gegg-Harrison, Carlson, and Tanenhaus (2013). Klein et al. (2013) were examining different types of referring expressions: regular definite noun-phrases, such as *"the concert"* and weak definite noun phrases, such as *"the hospital"*. Unlike regular noun-phrases, weak definite noun phrases do not need to uniquely refer. For example, one cannot say *John went to the concert and so did Sally* when they each went to different concerts; however, one can say *John went to the hospital and so did Sally* when they each go to different hospitals. In Klein et al. (2013)'s study, a magnet board was implicitly divided into two regions by painting one half of the board blue and the other half yellow. In an act-out task, John and Sally were on separate colors. After acting out *John went to the concert*, participants were reluctant to have Sally go to the same concert, if it meant crossing the implicit boundary. They noted that participants were reluctant to cross the boundary even when they were explicitly told that they could and were given practice and filler trials that required them to cross the boundary.

Here we directly test the hypothesis that abstract perceptual cues can structure the referential domains used in reference resolution with ambiguous referential terms. Given an ambiguous referential expression, listeners were asked to choose between two identical possible referents, one of which was perceptually grouped with a previously mentioned entity. To group referents on the basis of perceptual cues, we appealed to the Gestalt principles, such as proximity, common background, etc. (Wertheimer, 1938). In the current experiments, we ask three questions: First, will participants use perceptual grouping to create referential domains, or linguistic contexts? Second, if they do so, are these linguistic contexts defined by

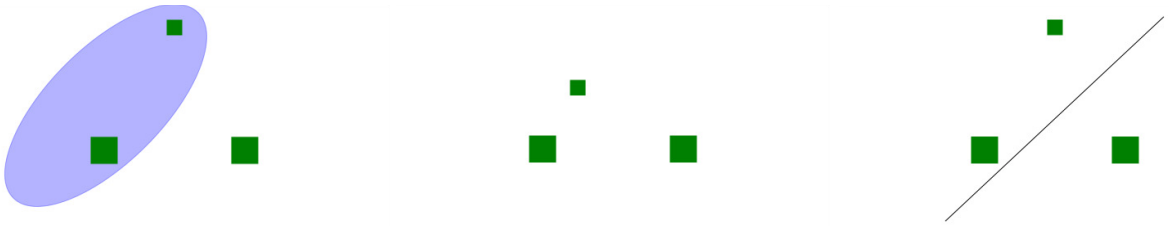

Figure 1: Stimuli used in Experiment One. Left to right: Common Background, Proximity, Boundary Region

the presence of or as a function of the perceptual cues? Third, do these referential domains really reflect perceptual grouping or can they be reduced to attentional factors, such as such as cognitive status (Gundel, Hedberg, & Zacharski, 1993), salience (Ariel, 1990), and attentional state (Gordon, Grosz, & Gilliom, 1993), which are often posited as primitives that determine referential domains?

## Experiment One

To investigate the role of perceptual cues in establishing a linguistic context, we presented participants with a visual array containing three squares, one smaller than the other two. The small square was perceptually grouped with one of the other squares. Participants were first instructed to click on the small square. If the perceptual cue is used to construct a linguistic context, we expect the linguistic context, or referential domain, will narrow to only include items in that perceptual grouping. We predicted that if the next referential expression is ambiguous between a perceptually grouped referent and an ungrouped referent, participants will resolve the ambiguity by selecting the referent within the narrowed context.

Given that linguistic contexts permit listeners to make rich pragmatic inferences, we predicted that the ambiguous referential expression would differentially influence reference resolution. If the ambiguous referential expression is non-specific toward any aspect of the object (e.g., *"the other one"*), we expected participants to select the object in the perceptually established context. On the other hand, if the next referential domain is over-informative (e.g., *"the big one"*), people will use the over-informativeness to pragmatically infer that the intended referent is outside of the narrowed referential domain.

### Methods

**Participants** 180 participants were recruited via Amazon Mechanical Turk and paid for their participation, which lasted approximately 3 minutes. All participants passed a sound check task to ensure their speaker system was functional.

**Materials and Design** Referential expression (*"the big one"* vs. *"the big square"* vs. *"the other one"*) was manipulated between subjects. Perceptual grouping factor (Common Background vs. Proximity vs. Region Boundary) was manipulated within subjects.

On each of the six critical trials, participants were presented with an image that contained three squares arranged on the vertices of an equilateral triangle (see Figure 1). The square on the top of the triangle was always slightly smaller than than the other two. On each critical trial, the small square was perceptually grouped with one of the bottom squares according to one of the following grouping factors: Common Background, Proximity[1] or Region Boundary. The orientation (right vs. left) of the grouping factor was counterbalanced across trials. An additional six filler trials were constructed to ensure that the participants were attending to the task. On these trials, a circle, a star and a triangle were randomly arranged along the vertices of the equilateral triangle. All of the instructions were recorded by a male speaker of Standard North American English.

**Procedure** The experiment was presented online using psiTurk (McDonnell et al., 2012). Before the experiment began, participants answered a demographic questionnaire and completed a sound check task (i.e. type what you hear). Sixty participants were randomly assigned to each of the referential expression conditions. On each trial, participants were presented with an image. On critical trials, participants heard the instructions, *"Click on the small square."* When participants clicked on an object in the image, their choice was recorded and the object disappeared. Participants were then instructed, *"Now click on..."* followed by the referential expression assigned to them. When the participant clicked on an object, their response was recorded. The image was then removed and the next trial began after 1 s. On filler trials, participants were instructed to first click on the circle and then click on the star. Their responses were recorded to ensure active participation in the task.

### Results and Discussion

All participants had accuracy scores on filler trials greater than 90%. Data for each referential expression were analyzed using a mixed effect logistic regression with subject and item intercepts. As seen in Figure 2, participants assigned

[1]For the Proximity manipulation, the small square was placed midway along the edge it would usually share with the square it is grouped with.

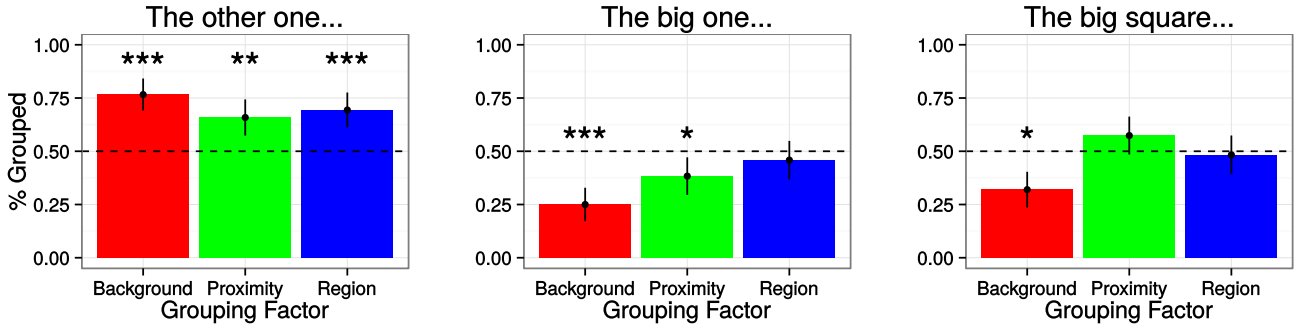


Figure 2: Results of Experiment One. Error lines represent 95% bootstrapped confidence intervals. Stars represent $p$ values comparing mean to chance, 50% ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

*"the other one"* as their ambiguous referential expression were more likely to choose the perceptually grouped square. This suggests that the perceptual grouping cues established a potential linguistic context and when placed in that context via the initial referential expression (i.e., *"Click on the small square"*), participants will remain in the context even though there is an identical referent visually available outside the context.

Participants assigned *"the big one"* as their ambiguous referential expression preferred the ungrouped square. One possible explanation for selecting the ungrouped square is that participants drew a pragmatic inference about the speaker's intended referent. If the speaker had intended to refer to the grouped object, there is no need to be over-informative. This is Grice's maxim of quantity (Grice, 1975). On the other hand, if a speaker was attempting to refer to the ungrouped object, they could flout the maxim by being over-informative.

Participants assigned *"the big square"* were at chance between choosing the perceptually grouped and the ungrouped square for the Proximity and Region Boundary grouping factor. One possible explanation for these results is that participants were torn between two interpretations of the referential expression. At face value, a participant entering the perceptually induced context might label the objects *"the small square"* and *"the big square."* In this case, the referential expression is unambiguous and the participant should select the grouped square. If the participant does not initially label both objects in the context, the referential expression might be considered over-informative and, thus, warrant the same pragmatic inference as seen with participants presented *"the big one."* In this scenario, the participants average response might reflect two different interpretations of the referential expression that cancel any effect out.

Taken together, these results are consistent with the notion that perceptual cues guide the establishment of context and further suggest that the inclusion of a non-informative contrast pushes the listener outside of a previously established linguistic context. Interestingly, these results were not completely consistent across grouping factors; participants assigned to *"the big square"* condition significantly preferred the ungrouped square when the perceptual grouping cue was Common Background and participants assigned to *"the big one"* condition did not show a preference when the perceptual grouping cue was Region Boundary. We hypothesized that these anomalies might be due to the varying magnitude of the grouping cues. Common Background is a strong grouping cue, which might have hindered participants from considering both squares as part of the referential domain. Compared to Common Background, Region Boundary might not have been a strong enough grouping factor to construct a linguistic context. As a first step following up on this interpretation, we tested whether the strength of a grouping cue influenced the construction of a linguistic context.

## Experiment Two

We disentangled whether the presence of a grouping cue is sufficient to establish a linguistic context or if the effectiveness of the grouping cue depends on its magnitude by presenting participants with trials manipulating the strength of the Region Boundary grouping cue. If the construction of a linguistic context depends on the magnitude of the grouping cue, we predict that as the magnitude increases, more participants will use the cue to create a context and resolve the ambiguous referential expression (i.e., *"the other one"*) by selecting the perceptually grouped item. Whereas if the construction of a linguistic context only requires the presence of a grouping factor, we expect that across all magnitudes participants will construct a linguistic context and resolve the ambiguity with the object in that context.

### Methods

**Participants** An additional 60 participants were recruited via Amazon Mechanical Turk and paid for their participation, which lasted approximately 3 minutes. All participants passed a sound check task to ensure their speaker system was functional.

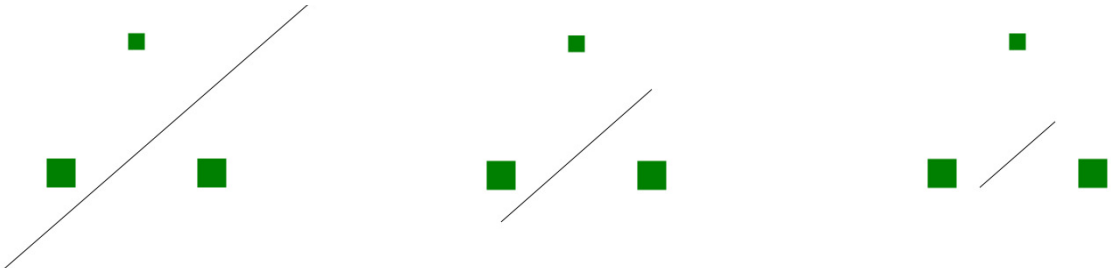

Figure 3: Stimuli used in Experiment Two.

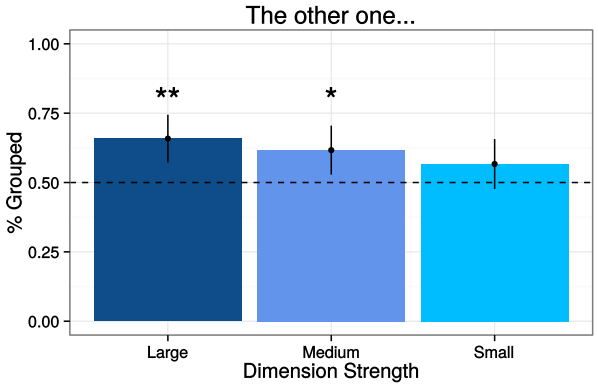


Figure 4: Results of Experiment Two. Error lines represent 95% bootstrapped confidence intervals. Stars represent $p$ values comparing mean to chance, 50% ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

**Materials** Similar to Experiment One experimental trials contained squares located on the vertices of an equilateral triangle. The top square was grouped with one of the otherwise identical bottom squares using the Region Boundary grouping cue at three different magnitudes (see Figure 3). At the strongest magnitude, the line separating the two squares bisects the screen. At medium strength, the line separating the two squares spans the length of the square encompassing the equilateral triangle. At the weakest magnitude, the line separating the two squares spans its length within the equilateral triangle. The orientation of the grouping factor (left vs. right) was counterbalanced across trials. The same filler trials as Experiment One were included to assess participant's attentiveness to the task. The same instructions from Experiment One were used.

**Procedure** The same procedure as Experiment One was used with one exception: all participants heard the ambiguous referential expression, *"the other one"*.

### Results and Discussion

All participants had accuracy scores on filler trials greater than 90%. Data were analyzed using a mixed effect logistic regression with subject and item intercepts. Replicating Experiment One, participants preferred to select the perceptually grouped square (see Figure 4). Moreover, as the strength of the perceptual cue decreased, participants were less likely to use the perceptual cue to establish a linguistic context, reflecting a gradient effect in subject averages. Future work will examine this effect in individuals.

## Experiment Three

Having now provided evidence for the role of perceptual cues in establishing linguistic context, it is important to distinguish the contribution of perceptual cues from the contribution of previously appealed to non-linguistic concepts such as salience (Ariel, 1990) or attentional state (Gordon et al., 1993). The previous appeals to salience nested within linguistic mention accounts do not focus on the narrowing of the referential domain but rather frame the problem as certain objects in the visual array attracting more attention as compared to the other objects in the array. One possible interpretation of the results of the previous two experiments is that the perceptual grouping factor increased the salience of the grouped object and participants selected the most salient object in the visual input. To demonstrate that our result is not an effect of salience, we conducted a variant of Experiment One in which the perceptually ungrouped square was cued before the trial. Cueing the location of the ungrouped square should increase its salience compared to the perceptually grouped square. Therefore, if participants resolve the ambiguity using perceptual cues, they should select the perceptually grouped object. Whereas, if participants resolve the ambiguity by selecting the most salient thing in the visual input, they should select the ungrouped object.

### Methods

**Participants** An additional 60 participants were recruited via Amazon Mechanical Turk and paid for their participation, which lasted approximately 5 minutes. All participants passed a sound check task to ensure their speaker system was functional.

**Materials and Design** The same images and recordings from Experiment One were used.

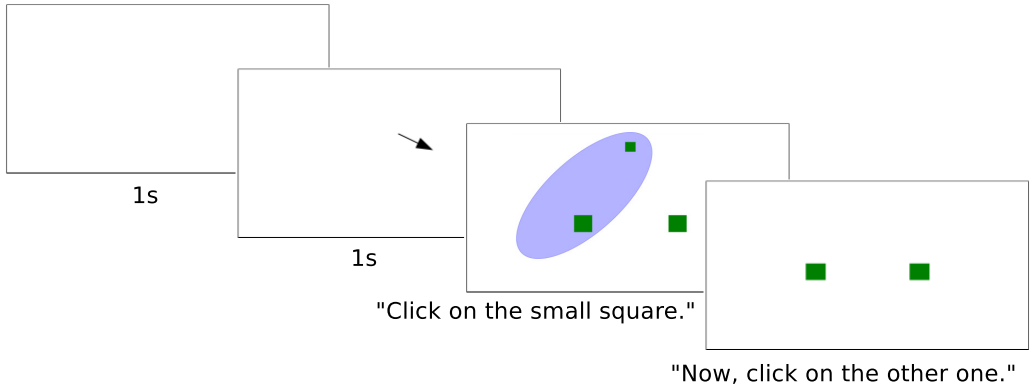


Figure 5: Sample trial from Experiment Three. All three grouping factors were used (only Common Background is illustrated).

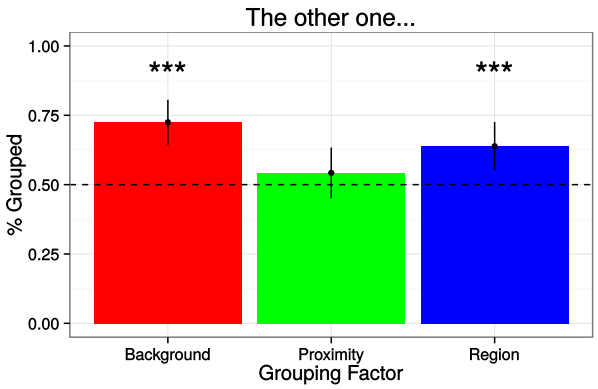


Figure 6: Results of Experiment Three. Error lines represent 95% bootstrapped confidence intervals. Stars represent $p$ values comparing mean to chance, 50% ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$)

**Procedure** The same procedure as Experiment One was adopted with two exceptions: 1) the only referential expression used was *"the other one"* and 2) before all experimental trials, an arrow pointing toward the location of the ungrouped item appeared in the middle of the equilateral triangle for 1 s (see Figure 5).

### Results and Discussion

All participants had accuracy scores on filler trials greater than 90%. Data were analyzed using a mixed effect logistic regression with subject and item intercepts. As can be seen in Figure 6, participants presented with the Common Background or Region Boundary grouping factor significantly preferred to resolve the ambiguity with the perceptually grouped object, suggesting the effect of perceptual cues on linguistic context is not an effect of salience[2]. Participants presented with the Proximity grouping factor were at chance between selecting the salient object and selecting the perceptually grouped object. In light of Experiment Two, one possible explanation for the lack of an effect is that the Proximity grouping factor was not strong enough to establish a linguistic context.

## General Discussion

As people listen to speech, they develop rich contextual representations that allow them to draw pragmatic inferences. In Experiment One, we showed that participants can use perceptual cues to establish a linguistic context. When moved into a linguistic context, they will resolve ambiguities locally unless the referential expression prompts the listener to pragmatically infer that the intended referent is not in the local context (via being over-informative). Experiment Two replicated the finding of Experiment One, that listeners use perceptual cues to establish linguistic contexts, and further demonstrated that a perceptual cue's ability to establish a linguistic context is gradient with respect to its magnitude. By explicitly drawing attention to objects outside the perceptual grouping, Experiment Three showed that the establishment of linguistic context was not due to the salience of perceptually grouped objects or attentional effects. These results serve as a first step toward developing a formal model for the establishment of linguistic context.

Ambiguity abounds in both printed language and in communicative encounters, yet we are remarkably quick at processing language and resolving ambiguities at all levels of processing. Explanations of how ambiguity is resolved often appeal to context without specifying what that context is and how it was established. Once we have a formal model of context construction, we can begin to explore the exact mechanisms of ambiguity resolution including how context and other components of online sentence processing (e.g., attentional factors) interact.

For example, recent research in discourse processing has disentangled the independent contributions of linguistic men-

[2]In an unpublished experiment, we attempted to group the squares with Common Motion. Participants were at chance in selecting grouped and ungrouped squares, which further suggests attentional salience does not influence the establishment of context.

tion and the presence of items in the visual array to the construction of the referential domain using the *contrastive adjective effect* (Sedivy, Tanenhaus, Chambers, & Carlson, 1999). When listeners hear a definite nominal phrase modified by a pre-nominal scalar adjective (e.g., *the big...*), they infer that the speaker is referring to something in the context, or referential domain, that varies along the adjective's contrast dimension. As a result, the referential domain narrows to entities within the contrastive set (i.e., the set of things in the context that vary along that dimension). One limitation of the original visual world experiments is that a contrastive set is always visually present. Therefore, results could not dissociate whether the narrowing of the referential domain was due to the visual context including a contrast set or simply a product of linguistic mention. By manipulating the availability of contrastive referents in the visual array, Wolter, Gorman, and Tanenhaus (2011) demonstrated that the narrowing of the referential domain is actually due to the linguistic mention alone. Further support for an independent effect of linguistic mention comes from work by Kim, Gunlogson, Tanenhaus, and Runner (2015) showing that focus operators introduce referential expressions to the referential domain.

While most research indirectly contributes to the definition of linguistic context by appealing to specific factors thought to be encompassed by the linguistic context, our work and the aforementioned work on linguistic mention demonstrates that by exploiting the ability of these factors to influence contexts, we can directly and empirically work toward a formal model of linguistic context. The key to this approach is to select an ambiguity such that the factor of interest is independent of the resolution of that ambiguity. Here we have demonstrated the approach with low level perceptual factors. Past research has applied this approach to linguistic mention. Future research should extend this approach to other factors relevant to the construction of linguistic contexts.

## Conclusion

In conclusion, listeners construct rich contextual representations that support pragmatic inferences using perceptual cues within the visual input. The effectiveness of the perceptual cues in establishing linguistic context depends on the magnitude of the cue, and the use of perceptual cues in establishing linguistic context is independent from the effects of salience in reference resolution. These results serve as a first step toward developing a formal model for the establishment of linguistic context by identifying the influence of perceptual cues.

## Acknowledgments

This research was supported by a NICHD grant HD27206 (MKT).

## References

Ariel, M. (1990). *Accessing noun-phrase antecedents (rle linguistics b: Grammar)*. Routledge.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of memory and language*, *47*(1), 30–49.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of experimental psychology: Learning, memory, and cognition*, *30*(3), 687.

Clark, H. H., & Marshall, C. (1978). Reference diaries. In *Proceedings of the 1978 workshop on theoretical issues in natural language processing* (pp. 57–63).

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science*, *17*(3), 311–347.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics.* New York: Academic Press.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *language*, 274–307.

Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition*, *139*, 28–49.

Klein, N. M., Gegg-Harrison, W. M., Carlson, G. N., & Tanenhaus, M. K. (2013). Experimental investigations of weak definite and weak indefinite noun phrases. *Cognition*, *128*(2), 187–213.

McDonnell, J., Martin, J., Markant, D., Coenen, A., Rich, A., & Gureckis, T. (2012). *psiturk (version 1.02)[software]. new york, ny: New york university.*

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147.

Wertheimer, M. (1938). Laws of organization in perceptual forms. *A source book of Gestalt psychology*, 71–88.

Wolter, L., Gorman, K. S., & Tanenhaus, M. K. (2011). Scalar reference, contrast and discourse: Separating effects of linguistic discourse from availability of the referent. *Journal of memory and language*, *65*(3), 299–317.