

Good Enough Language Processing: A Satisficing Approach

Fernanda Ferreira (fernanda.ferreira@ed.ac.uk)

Paul E. Engelhardt (Paul.Engelhardt@ed.ac.uk)

Manon W. Jones (manon.wyn.jones@ed.ac.uk)

Department of Psychology, University of Edinburgh
Edinburgh, UK EH8 9JZ

Abstract

The “Good Enough” approach to language comprehension assumes that listeners do not always engage in full detailed processing of linguistic input. Rather, the system has a tendency to develop shallow and superficial representations when confronted with some difficulty. In this study, we investigated Good Enough processing using a challenging version of the Visual World Paradigm, in which participants had to process a garden-path sentence, and then complete a second instruction. The main hypothesis was that if there is a tradeoff between dealing with current difficulty and keeping up with the incoming material, then we might expect superficial processing of the garden path, which would be observed in more errors. The results showed more errors in completing the garden path instruction compared to the immediately following second instruction. In this version of the task, participants did not show evidence of using the visual context to constrain the interpretation of a temporary ambiguity.

Keywords: Visual World Paradigm, Good Enough processing, syntactic ambiguity resolution, satisficing, eye movements.

Introduction

One prominent theory that has been offered to account for the fact that listeners will often develop inaccurate representations is called “Good Enough” processing (Ferreira, Ferraro, & Bailey, 2002; Ferreira, & Patson, 2007; Sanford & Sturt, 2002). Under this explanation, listeners may generate an interpretation of an ambiguous or a temporarily ambiguous utterance that is not consistent with the actual input. Instead the system has a tendency to generate shallow or superficial representations, and much of the time the inaccuracies are consistent with the plausibility of events in the real world (Christianson, Hollingworth, Halliwell, & Ferreira, 2003; Ferreira, Christianson, & Hollingworth, 2001). For example, Christianson et al. found that subjects would often misinterpret sentences like examples (1) and (2) below, in which they believed that the italicized noun phrase was simultaneously the object of the first verb and the subject of the second verb. This interpretation is not consistent with a grammatical syntactic structure (see also Engelhardt, Patsenko, & Ferreira, 2007, for similar results with spoken versions).

(1) While Anna bathed *the baby* played in the crib.

(2) While the man hunted *the deer* ran through the woods.

The theory of Good Enough processing focuses on two main issues. The first is that representations formed from complex or difficult material are often shallow and incomplete. The second is that limited information sources are often consulted in cases where the comprehension system encounters difficulty. It is important to note that the comprehension errors that have been observed in previous studies are not completely wrong. Rather the system seems to make systematic errors. In the subject-object ambiguity in example (2), it is possible that the man was hunting deer, but crucially this inference is not specified in the sentence.

In previous work, we have argued that this result is evidence that language comprehension is a matter of satisficing. Satisficing is a term originally used by Herbert Simon to describe the search for information in order to make a decision (Simon, 1956). He assumed that decision makers would search for information until a solution was found that achieved or surpassed aspiration levels. A great deal of work in the field of decision making has focused on the use of heuristics in a wide variety of tasks (Tversky & Kahneman, 1974). One of the overarching ideas from that literature is that humans in many instances do not consider all available information when making a decision. This is especially true in situations when time and cognitive resources are limited. What is surprising about the results from many studies is that people can perform quite accurately, though not perfectly, by applying a small set of simple heuristics (Gigerenzer, 2008; Gigerenzer & Selten, 2001). The main characteristic of these heuristics is that they operate quickly, using minimum processing resources.

Taking this perspective as background, some interesting similarities emerge with models of sentence comprehension. For example, the Garden Path model of sentence processing assumes that the comprehension system uses simple heuristics for building initial interpretations (Frazier, 1987). The most important of these is Minimal Attachment, which assumes that the parser will always build the simplest syntactic structure first. The competing approach in sentence processing, and one that has had a great deal of support over the past twenty years involves an interactive processing architecture (MacDonald, Pearlmutter, & Seidenberg, 1994). The key aspect of this type of model is that the parser can immediately draw on any relevant

information source. Some models even assume that all possible interpretations of an ambiguity will be maintained in parallel until one is selected (Tanenhaus, & Trueswell, 1995). This type of processing is akin to what researchers in decision making refer to as “unbounded rationality”.

If language comprehension did not depend on processing resources, then it is hard to imagine why people would ever make mistakes at all, let alone the seemingly systematic misinterpretations that have been observed with both garden path sentences and passives (Ferreira, 2003). We hypothesize that the good-enough representations that have been observed in previous work arise by the application of heuristics, which allow the processor to operate quickly while minimizing the demand on cognitive resources. Indeed, much of the evidence for Good Enough processing can be explained by a semantic plausibility heuristic. If the syntactic parse breaks down, then the system will derive an interpretation most consistent with what is likely to have occurred given the normal state of affairs in the real world.

In the current work, however, we were interested in exploring the situations under which people will abandon the revision of a garden path sentence. In spoken language, listeners do not control the rate at which they receive input. Therefore, if listeners encounter difficulty, they do not have the luxury of taking unlimited processing time because new material is likely to be arriving. In these cases, the processing system may have a tendency to generate a good-enough interpretation of the difficult part in order to keep up with the new material. Alternatively, the system might continue to work on the difficulty and sacrifice processing of the new input. Examining these sorts of tradeoffs and the conditions under which people engage in Good Enough processing were the goals of the current study. An additional goal was to investigate whether the processing system, under challenging conditions, will draw on the types of information sources that have been argued to support interactive processing architectures.

Experiment

In the active version of the Visual World Paradigm, participants are presented with a 2 x 2 grid of objects, and they are given spoken instructions to move objects in a workspace (Ferreira & Tanenhaus, 2007). Two quadrants, typically on the left side of the display, contain the target and the distractor object, and are usually the objects that are moved first. The right side of the display contains two goal locations. Participants hear either ambiguous or unambiguous instructions featuring a prepositional phrase modifier. The commonly reported result is that when subjects hear an instruction such as *put the apple on the towel in the box*, and the display contains a single apple, participants will often fixate the empty towel shortly after hearing the ambiguous prepositional phrase (Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). These fixations are interpreted as showing that subjects momentarily

entertained the goal analysis of *on the towel*. When the display contains two apples, a different fixation pattern is reported. With two-referent displays, subjects almost never look to the empty towel, but instead show competition between the two apples after which they look to the correct goal (i.e. *the box*). This fixation pattern is taken as evidence that the visual context (i.e., the presence of two apples) can immediately resolve the temporary ambiguity, and is evidence for an interactive processing architecture.

In the current studies, we addressed two main questions. The first was to examine Good Enough processing. To do this, we created a situation to specifically examine the tradeoffs between local difficulty and later processing. The second question we asked was whether subjects show evidence of using visual context to resolve a temporary ambiguity when visual displays are more demanding than in previous visual world studies.

To increase the processing demands we made three modifications. The first was that we used arrays of twelve objects instead of four. Second, we eliminated the preview. In this study, the objects appeared at the same time as the spoken instruction began. These two manipulations affected visual demand. In addition, we added a second instruction, so that subjects could not linger on the garden path sentence, without sacrificing processing of the subsequent words. A two-referent example display is shown in Figure 1, and the corresponding instructions are given in (3) and (4). A one-referent display was similar, but the lone book was replaced with, for example, a football. The tradeoffs that we predicted could be observed in several measures. Do subjects put the correct book on the chair or in the bucket? Do subjects show evidence of considering the empty chair as a goal location? How quickly do subjects execute the instructions?



Figure 1: Example display for the two-referent condition.

- (3) Put the book on the chair in the bucket. Then click on the balloon.
 (4) Put the book that’s on the chair in the bucket. Then click on the balloon.

Methods

Participants

Twenty-eight students from the University of Edinburgh participated. Participants were native speakers of British English, and had normal or corrected-to-normal vision.

Materials

Each visual display consisted of a 4 x 3 array of common objects taken from the Hemera Photo Objects database (Figure 1 is reduced color & quality). Each display was accompanied by a pair of auditory instructions. These utterances were recorded by a female native speaker of British English. The first instruction required the participant to move one object in the array to a specific location, and this instruction was either ambiguous or unambiguous. The ambiguous utterance was created by digitally excising the complementizer *that's* from the unambiguous instruction. The second instruction occurred immediately after the first and required the participant to click on a different object in the display. The visual array of objects on average subtended 22° degrees of visual angle horizontally and 17° vertically, for a viewing distance of 60 cm. For each of the 24 critical displays, both one-referent and two-referent versions were created. These were placed into four lists that were counterbalanced with instruction type. Lists were rotated in a Latin Square design, so that each subject saw each display only once.

Design and Procedure

We used a 2 x 2 mixed design. Display type contained either one or two referents and was manipulated between subjects. Instruction type was either ambiguous or unambiguous, and was manipulated within subjects. Participants completed six practice trials, 24 experimental trials, and 96 fillers. Participants were told that they would see an array of pictures, which would be accompanied by instructions to perform certain actions on the objects depicted. They were asked to execute the instructions as quickly and accurately as possible. The session lasted approximately 35 minutes.

Results

Results were analyzed using a logit mixed effects model because our dependent variables were binomial (Baayen, Davidson, & Bates, 2008). The first dependent variable was whether participants made a movement or click error (see “total correct” in Table 1). The second dependent variable was whether a fixation was made to a particular object during a specific time window. We analyzed four 1.5 sec time windows, which were time locked to the onset of each of the nouns in the instructions. Each time window was shifted forward by 200 ms to account for saccade planning time (there was almost no overlap between any of the four windows). Logit mixed models have been advocated as more appropriate for binomial data than are ANOVAs

(Jaeger, 2008). These models allow multiple random factors, so we included both subjects and items.

For each analysis, we first created a baseline model, which included an intercept and the two random factors. This model was then compared to successive models that included an additional predictor. If the inclusion of an additional factor significantly improves fit over the baseline, then we can conclude that it accounts for a significant amount of variation in the data. Factors were entered one after the other, and a third model tested the interaction. We assessed model improvement via log-likelihood ratio tests using the R statistical package (Bates, Maechler, & Dai, 2008). This test compares models using a χ^2 test which determines whether an additional predictor significantly improves model fit. Each model includes an intercept and slopes representing the effects of each of the factors in the model. In cases where a predictor significantly improved fit, the Wald statistic was used to show that coefficients differed significantly from zero (Agresti, 2002).

Errors

We began the analysis by examining the number and types of errors participants made (see Table 1). As can be seen from the table, there are substantially more errors with the two-referent display than with the one-referent display. Trueswell et al. (1999), and Farmer, et al. (2007) both reported a main effect of instruction type, in which there were more errors with the ambiguous instructions. However, in both studies there were more errors in the one-referent compared to the two-referent condition. Our results showed the opposite effect of number of referents. The analysis showed that model fit was significantly improved with the inclusion of both instruction type and display type as factors [$\chi^2(1) = 17.46, p < .001$]. However, fit did not improve when the interaction was included [$\chi^2(1) = 2.04, p = .15$].

Table 1: Summary of mouse movements for each condition

	<u>Ambiguous</u>		<u>Unambiguous</u>	
	1-ref	2-ref	1-ref	2-ref
Total correct	95%	69%	97%	89%
Distractor pickup	2	42	0	9
Distractor drop	1	23	0	3
Click error	2	9	2	5

To evaluate tradeoffs between the garden path and the second instruction, we examined three types of errors. “Distractor pickups” were errors in which the subject picked up the single book. “Distractor drops” were when a subject placed a book on the incorrect goal (i.e. the chair). Lastly, “click errors” were when the subject clicked on the wrong object. The majority of errors were when the participant picked up the distractor, and on about half of those trials the participant dropped it on the incorrect goal. In contrast, there were relatively few click errors.

Eye Movements

For the eye movement analysis, we began by examining the time window beginning with the onset of the first noun (e.g. book). Here we were particularly interested in examining both looks to the target object (e.g. the book on the chair) and to the distractor (e.g. the single book or the football). Figure 2 shows the proportion of trials with a fixation to the target. Error bars show the standard error of the mean.

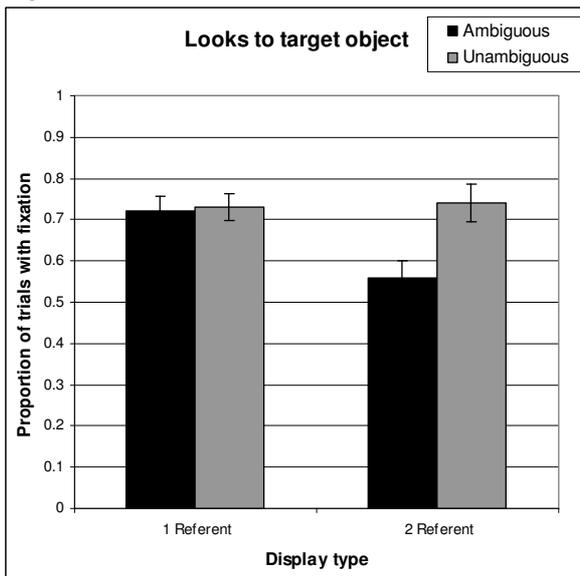


Figure 2: Proportion of trials with fixation to target object.

The mixed model analysis showed that model fit was significantly improved with the inclusion of the interaction [$\chi^2(1) = 2.68, p < .01$]. The interaction is primarily driven by the decreased likelihood of fixating the target object in the two-referent ambiguous condition. The analysis of looks to the distractor object showed the expected effect of display type, with the one-referent display showing ~30% of trials with a fixation, and the two-referent display showing ~70%. There were no differences based on instruction type and no interaction. The model containing only display type was a significantly better fit over the baseline model [$\chi^2(1) = 43.2, p < .0001$]. This analysis shows the expected pattern. In the two-referent condition there should be competition between the two potential referents (e.g. book's), and in the one-referent condition the distractor object is an irrelevant item, and so looks to this object should be low.

The second time window was time locked to the onset of the second noun in the instruction (e.g. chair). Again, we analyzed a 1.5 sec window, but this time we were interested in examining the looks to the incorrect goal, which in the example display is the empty chair. The results from this analysis showed that model fit was only improved with the inclusion of display type [$\chi^2(1) = 3.81, p = .05$]. The proportion of trials with a fixation to the incorrect goal are shown in Figure 3, and there is a trend towards an

interaction. The interesting result from this analysis is that it is opposite of previous studies. The “typical” pattern is for there to be more looks to the incorrect goal in the one-referent ambiguous condition, but here we see significantly more looks with two-referent display. We will discuss possible reasons for the conflicting data in the Discussion.

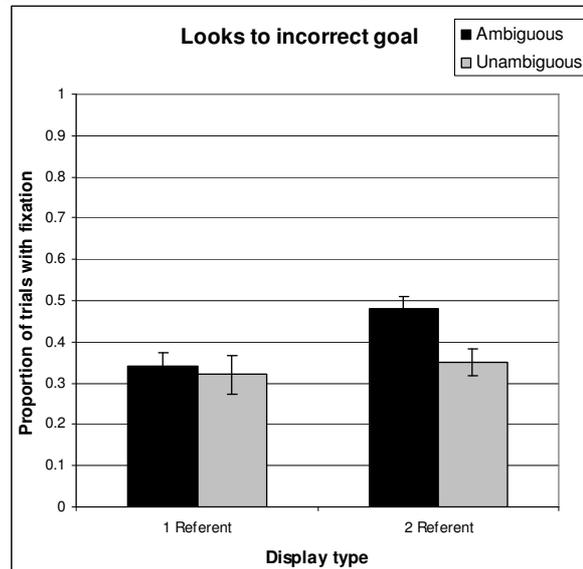


Figure 3: Proportion of trials with fixation to incorrect goal.

The analysis of the third and fourth time windows showed very similar results (see Figure 4). The third window was time locked to the onset of third noun (e.g. bucket). The results showed that model fit was improved only with the inclusion of instruction type [$\chi^2(1) = 9.63, p < .01$]. The fourth time window was time locked to the onset of the click object (e.g. balloon), and it showed virtually the same result. Model fit was improved with the inclusion of instruction type [$\chi^2(1) = 10.1, p < .01$]. In these two windows, we see a large effect of instruction type in which the ambiguous instructions showed fewer looks than did the unambiguous instructions. However, in all conditions the majority of trials contained a fixation to the relevant object.

Discussion

This study had two goals. The first was to examine whether there is a tradeoff between current processing difficulty and the comprehension of subsequent material. The results showed substantially more errors on the garden path sentence than on the second instruction. This suggests that subjects have a tendency to sacrifice reanalysis of the garden path in order to keep up with the later material. This pattern of results is consistent with the assumptions of the Good Enough theory of language processing, which assumes that processing resources are limited, and therefore predicts that garden-path reanalysis processes will be curtailed if upcoming material must also be processed.

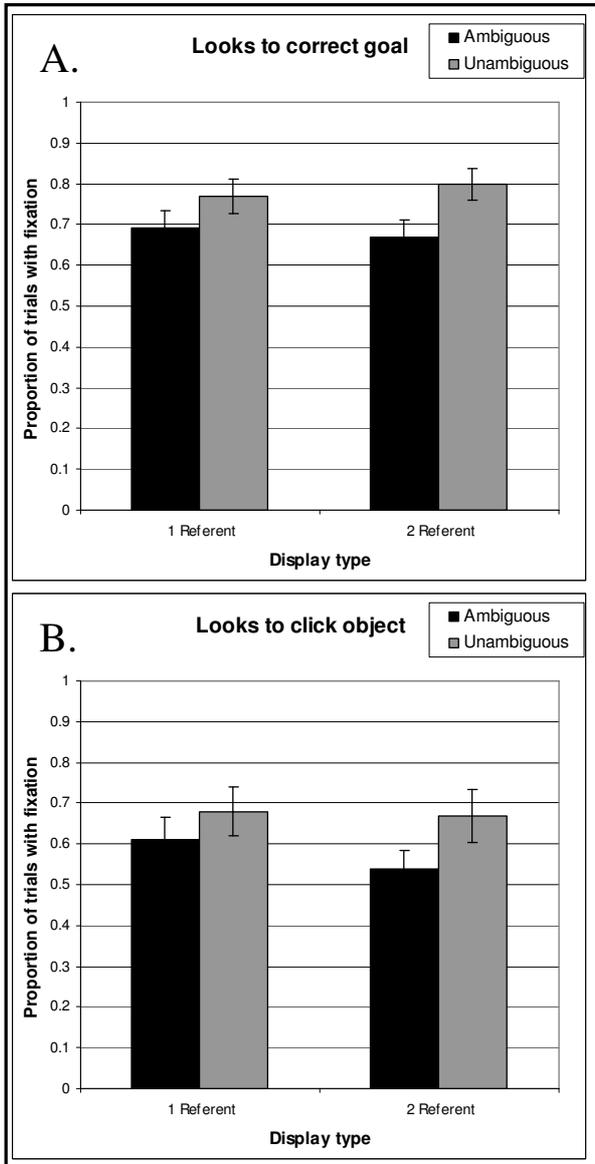


Figure 4: Panel (A) show proportion of trials with fixation to correct goal. Panel (B) shows proportion to click object.

The second goal of the study was to examine the use of visual information on the resolution of temporary syntactic ambiguity. Here the results showed a different pattern from that reported in previous visual world studies. The first unique finding was that there were few looks, essentially at chance, to the incorrect goal with the one-referent display. Recall that looks to the incorrect goal in response to hearing the ambiguous prepositional phrase with the one-referent display were previously interpreted as evidence that subjects initially adopted the goal analysis of the ambiguous phrase.

We believe that the absence of looks to the incorrect goal in this condition can be explained by our previous work in

which we examined the types of utterances that people produced with a one-referent display using the standard four object visual arrays (Engelhardt, Bailey, & Ferreira, 2006). Those results showed that on one-third of trials, naïve subjects produced an unnecessary modifier when instructing another person to move the target object to the correct goal. For example, one third of trials were *Put the book on the chair in the bucket*, even though *Put the book in the bucket* would have been communicatively sufficient. We also showed that when subjects had to produce an instruction to move, for example, a book on a chair to an empty chair, almost all of the instructions contained a pre-nominal modifier to distinguish the two chairs (e.g. *Put the book on the other chair.*). Therefore, “unnecessary” modifiers are relatively common, and because the noun *chair* isn’t modified with a word such as *other*, subjects should interpret the prepositional phrase as a modifier rather than a goal even in one-referent displays.

The second novel finding in this study was that we found looks to the incorrect goal in two-referent contexts. Recall that Tanenhaus et al. (1995) showed few looks to the incorrect goal, and they interpreted this result as evidence that visual context can immediately influence the resolution of the temporary ambiguity. In contrast, we did get significantly more fixations to the incorrect goal with the two-referent display. The interaction for this particular analysis did not quite reach significance (it is possible the study is underpowered with only 14 subjects in each group), but there are clearly more looks with the ambiguous instruction. Setting the significance issue aside for the moment, it is still surprising that we observed such a high number of trials with a fixation to the incorrect goal with the two-referent display, because this is exactly the opposite of what has been found in previous studies. What these results show is that with this version of the task, subjects do not seem to immediately adopt the modifier interpretation. Therefore, it does not seem that the subjects use visual context to help interpret the ambiguous linguistic input when task demands are high.

One issue that was brought up in the review process was that since we eliminated the preview phase and increased the number of distractors, subjects perhaps did not have enough time to process the display, thus making the two-referent display a one-referent display. We believe that this explanation can be ruled out because looks to the target object in the two-referent ambiguous condition are well over chance, and looks to the distractor object were only slightly greater (.68 vs. .56). This suggests that on the majority of trials participants did fixate the target object during the first time window. Moreover, the error movement analysis showed that on 25% of trials participants began moving the distractor object, and on over half of those trials they actually made the garden path error. What is clear from these results is that subjects have tendency to get the goal interpretation despite the fact that the majority of trials contained a fixation to both the target and the distractor. We

are currently running follow up experiments to isolate which of the unique features of our experimental set-up might have caused the distinct pattern of results.

In conclusion, we have shown that task demands affect processing of temporarily ambiguous sentences in complex visual contexts. We believe that the differences are likely due to the fact that previous paradigms were atypically easy, which perhaps allowed subjects to anticipate (or predict) the type of structure they might receive. The results also indicate that when the task is made more difficult, the two-referent condition becomes more difficult than the one-referent condition. We also showed that participants did not use the visual context to resolve the ambiguity in this particular situation. These results are broadly consistent with many results that have been obtained in the decision making literature (Goldstein & Gigerenzer, 1996), and further evidence that humans have a tendency to engage in good-enough processing.

Author Note

The authors would like to thank Fiona Allen, Oliver Stewart, and Laura Speed for the running experiment and the preparing stimuli. We would also like to thank Jens Apel and Martin Corley for help with the statistical analysis.

References

- Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 413-425.
- Bates, D., Maechler, M., & Dai, B. (2008). *lme4: Linear mixed-effects models using Eigen and Eigen++* [Computer software manual]. Available from <http://lme4.r-forge.r-project.org/>.
- Christianson, K., Hollingworth, A., Halliwell, J., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42, 368-407.
- Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean maxim of quantity. *Journal of Memory and Language*, 54, 554-573.
- Engelhardt, P. E., Patsenko, E., G., & Ferreira, F. (2007). Pupillometric indices of visual and prosodic information on spoken language comprehension. In D.S McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 971-976). Mahwah, NJ: Erlbaum.
- Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, 47, 164-203.
- Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: implications for models of reanalysis. *Journal of Psycholinguistic Research*, 30, 3-20.
- Ferreira, F., Ferraro, V., & Bailey, K. G. D. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.
- Ferreira, F., & Patson, N. (2007). The good enough approach to language comprehension. *Language and Linguistics Compass*, 1, 71-83.
- Ferreira, F. & Tanenhaus, M. K. (2007). Introduction to the special issue on language-vision interactions. *Journal of Memory and Language*, 57, 455-459.
- Farmer, T. A., Cargill, S. A., Hindy, N.C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science*, 31, 889-909.
- Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 601-681). Hillsdale, NJ: Erlbaum.
- Gigerenzer, G. (2008). Why heuristics work. *Perspectives on Psychological Science*, 3, 20-29.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.
- Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox*. Cambridge, MA: MIT Press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434-446.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review*, 101, 676-703.
- Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Science*, 6, 382-386.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129-138.
- Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447-481.
- 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.
- Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication* (pp. 217-262). San Diego: Academic Press.
- Trueswell, J. C., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence comprehension in young children. *Cognition*, 73, 89-134.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185, 1124-1131.