Tit-for-Tat: Effects of Feedback and Speaker Reliability on Listener Comprehension Effort

Nicole Craycraft* (ncraycraft@gmail.com)
Department of Psychology, University of Illinois, Urbana-Champaign

Zoe Kriegel* (zkriegel@kent.edu)
School of Health Sciences, Kent State University

Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Brain & Cognitive Sciences, University of Rochester

Jennifer Roche (jroche3@kent.edu)
School of Health Sciences, Kent State University

Abstract

Miscommunication is often seen as a detrimental aspect of human communication. However, miscommunication can differ in cause as well as severity. What distinguishes a miscommunication where conversation partners continue to put forth the effort from miscommunication where conversation partners simply give up? In this eye-tracking study, participants heard globally ambiguous statements that were either a result of an experimental error or speaker underspecification; participants either received positive or negative feedback on these ambiguous trials. We found that negative feedback, paired with the reliability of the message, will impact the amount of processing effort a comprehender puts forth—specifically, listeners were less forgiving of errors when they were penalized and when speakers' instructions lacked effort. This suggests that language users weigh conversational contexts and outcomes as well as linguistic content during communication.

Keywords: ambiguity; intentions; communication; comprehension; context

Introduction

Communication ideally ensures a successful exchange of information between speakers and listeners. Unfortunately, communication rarely functions ideally: a successful exchange will be riddled with unsuccessful attempts. But are all miscommunications equally disruptive? The answer to this question has implications for the mechanisms used in language processing.

In an egocentric account, speakers and listeners automatically use their own perspective, often failing to take the perspective of others in the conversation. Evidence for this view includes the finding that speakers are *more* likely to inadvertently reveal the identity of a hidden object when explicitly instructed to keep it secret, presumably due to the increased attention paid to it (Lane & Ferreira, 2008) and that speakers in a matching task are better able to tailor utterances to the past experience of a particular listener when item category labels are distinct for each listener rather than overlapping, reducing the memory load of pairing a specific item to a specific listener (Horton & Gerrig, 2005). In this view, miscommunication is a necessary consequence of egocentric thinking, and would often go undetected. Indeed, when task partners have different goals and were either uninformed or misinformed that their goals were the same, statements by both parties were routinely misinterpreted and only 1 pair out of 31 discovered their goals were different (Russell & Schober, 1999)

Contrasting this is an audience design account, where the perspective of and previous experiences with a conversational partner constrain interpretation or production. An assumption of this account is all parties in a conversation possess cooperative intent. Grice (1975) presented a theory of general maxims that communication must follow to be successful, including "make your contribution as informative as is required (for the purposes of the exchange)" as a maxim of Quantity and "do not say that for which you lack adequate evidence" as a maxim of Quality. From these maxims as well as others, we can derive that listeners assume speakers will provide true information with as much detail as a context necessitates. There is a body of literature suggesting that listeners can flexibly incorporate contextual information to intuit the intentions of speakers. Listeners use an ambiguous utterance as a signal of mismatching perspectives and readily adapt to the speaker's perspective to quickly identify a referent (Hanna, Tanenhaus, & Trueswell, 2003). Shared experiences may form a *referential pact* by establishing a potentially ambiguous phrase as a term for a referent; when a speaker exerts more effort and uses a new term for a previously mentioned referent,

* Indicates shared first authorship

it delays comprehension of the listener (Metzing & Brennan, 2003).

In this audience design framework, miscommunication would be a mismatch between perceived context and actual context. Roche, Paxton, Ibarra, and Tanenhaus (*under review*) posit that miscommunication can promote deeper processing and more care to resolve misaligned representations of the world. Broadly, most miscommunications can be explained as either an accident by a still cooperative party or as an indication of a partner who is an unreliable communicator.

Listeners suspend their expectations of correct usage when a demonstrably unreliable speaker mislabels common objects and misuses modifiers (Grodner & Sedivy, 2011). Even young children are sensitive to speaker reliability, as they usually rely upon adults over their peers for accuracy of new information. In fact, when preschoolers detected inaccuracies from an adult, they quickly overrode this tendency (Jaswal & Neely, 2006). In addition, young children also show capabilities to learn selectively from more reliable sources without specifically discussing prior reliability of relevant speakers (Birch, Vauthier, & Bloom, 2008). Speakers that continually provide false or vague information may cause listeners to disregard subsequent utterances by the same speaker. Speakers also have assumptions about listeners: namely, that they will pay attention. In Kuhlen and Brennan (2010), speakers told jokes they had previously learned. When speakers expected attentive listeners and received distracted listeners, they used fewer details.

In the following task, we tracked listeners' eye movements while they heard unambiguous or globally ambiguous (i.e., unresolvable) descriptions from a speaker during a matching task. Sometimes, the globally ambiguous description is due to a perspective mismatch; in other cases, the speaker simply did not provide enough information. During the globally ambiguous trials, half of the listeners received negative feedback and half received positive feedback. If miscommunication is processed purely egocentrically, trials with negative feedback should cause the participant to disengage from the task regardless of the source of the error. However, if listeners are sensitive not just to the *presence* of a miscommunication but the *cause* of it, the accidental errors should be more easily "forgiven" than lazy speaker errors, especially in the negative feedback cases. We expect listeners to process ambiguity differentially depending on the reason for the ambiguity and whether the ambiguity prevents communicative success.

We hypothesize that listeners understand communication as a shared experience and should process errors differently depending on the intention and outcome. We predict that listeners should exert more effort in processing the speaker's perspective if the error was the fault of the experimenter compared to errors due to speaker laziness. The consequence of the error should also affect the listener. If the listener is penalized for the speaker's laziness, then the listener should stop *working hard* too (tit-for-tat).

Methods

In the following study, we used a pseudo-confederate design (description below), which involved the ruse that the listener was interacting with a live person (similar to Roche, Dale, & Kreuz, 2010). During the task, the listener heard an instruction (pre-recorded statement) describing which object to click. The listener then saw feedback about the correctness of the choice she made. On some of the trials, the listener had difficulty making the correct decision because the speaker produced a globally ambiguous statement. The listener then learned the source of the ambiguity, which we predict will differentially affect processing effort on decoding the error and future effort.

Participants

Sixteen undergraduates from Kent State University (15 females; mean age = 21.5 years) participated for extra credit in a Speech Pathology & Audiology course, resulting in a total of 4,480 data points across the experimental trials. All participants were native speakers of American English with normal to normal-corrected vision. None reported speech or hearing impairments.

Materials & Stimuli

A 21inch iMac (experiment computer), Eyelink 1000 (eye-tracking computer), noise cancelling headphones, wireless mouse, and usb microphone (for pseudo-confederate recordings) were used. All materials were set up in a sound attenuated booth, with two chairs – one for the participant and one for the experimenter to control the eye tracking computer.

Visual stimuli. Participants were presented 4 shapes (square, circle, triangle, and star) x 4 colors (purple, blue, green, and red) x 2 sizes (big and small), resulting in a total of 32 possible shapes. On a given trial, two shapes were paired with each other, resulting in visual stimulus pairs that overlapped on zero, one, or two features.

Auditory stimuli. Participants were presented with pre-recorded statements from a Caucasian female talker that referenced one (e.g., big shape), two (e.g., big red shape), or three (e.g., big red triangle) possible features of the visual stimulus. The recordings were equated for RMS amplitude to adjust stimulus sound level for more comfortable listening. This process "turned up" the volume on the sound files that were at lower amplitudes to match the highest amplitude and did not affect recording quality.

Design & Procedure

An individual listener interacted with a confederate during the informed consent portion of the study, but the listener and confederate were separated during the experimental task. After consent was obtained, the listener was seated in the sound booth with the door open, and the confederate sat at a computer and microphone around the corner. The listener and confederate received separate instructions, but the listener could hear the experimenter providing the confederate's instructions. The listener and confederate were told they would interact with each other via a computer network – much like a cell phone conversation. In fact, the listener never interacted with the confederate during trials but, instead, heard pre-recorded statements by a pseudo-confederate.

Once the pseudo-confederate ruse was established, the listener was calibrated, validated, and drift corrected using the prescribed procedure for the Eyelink 1000. During the task, the listener viewed two objects on the computer screen in addition to a bullseye (see Fig. 1, *left panel*). By clicking the bullseye, the listener initiated the trial and received a pseudo-confederate instruction.

On any given trial, the instruction could have been ambiguous or unambiguous. Sometimes, the global ambiguity forced the listener to guess which object the speaker intended (e.g., "Click on the red shape," when paired with a visual display containing two red shapes; see Fig. 1, *left panel*). On the next screen, the listener received feedback indicating whether the choice was correct or incorrect. The content of the speaker's screen was also on this screen.

A 2 **Feedback** (*Consequence* vs. *No Consequence*; between-subjects) by 3 **Error Type** (*No Error*, *Experimenter Error*, or *Lazy Speaker Error*; within-subjects) design was used. The **Feedback** conditions determined the type of feedback listeners received on globally ambiguous trials. Listeners in the *Consequence* condition always received negative feedback on the globally ambiguous trials – indicating they chose the incorrect object (see Fig. 1, *right panel*). Listeners in the *No Consequence* condition always received positive feedback on the globally ambiguous trials – indicating they chose the correct object (green check mark, instead of a red X, as indicated in Fig. 1, *right panel*).

Feedback was crossed with **Error Type**. In the *Experimenter Error* condition, the speaker did not see the same objects as the listener – the source of the error was due to mismatching visual referents (see Figure 1, *right panel*). In the *Lazy Speaker Error* condition, the speaker accidentally described the overlapping feature instead of the disambiguating feature – the speaker was being lazy and not paying attention to detail. Trials without errors comprised the No Error condition.

Over the course of the experiment, each listener was presented with a total of 280 experimental trials: ~20% of trials included a global ambiguity (60 trials total). The global ambiguity trials were pseudo-randomly distributed across the experiment with the first error occurring at trial 57 (i.e., 912 critical data trials across all participants). As a reminder, these types of errors required the listener to make a best guess. The other 80% of trials included a *resolvable ambiguity* or *unambiguous* statement that did not affect the listener's ability to choose correctly.

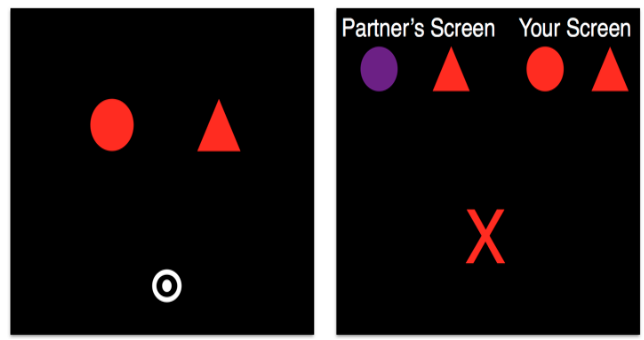

Figure 1. Sample experiment screen for an Experimenter Error trial, with negative feedback, when the listener heard: "*Click on the red shape.*" The *left panel* represents the Response/Instruction screen; the *right panel* represents the Feedback Screen.

Measures

Over the course of the experimental trial, dwell times were evaluated (i.e., how long the listener fixated within an interest area). Dwell time has been suggested to be a good measure of cognitive processing, and we use it to determine processing effort (Ehrlich & Rayner, 1981; Rayner & Duffy, 1986). It is expected that under moments of ambiguity that the listener should experience increased cognitive load while processing the ambiguity. However, interest here laid in the amount of time *and* effort the listener was willing to put forth understand the locus of ambiguity. Therefore, we chose to use dwell time measure, instead of pupil dilation – which is also an established measure of cognitive load (c.f., Kahneman, 1973).

The three areas of interest included the two objects on the Instruction Screen (Fig. 1, *left panel*) and the location of the "Partner Screen" information on the Feedback Screen (Fig. 1, *right panel*). Dwell times were calculated using the Eyelink Dataviewer software based on a fixation that landed in the interest area for a predetermined algorithm that calculates the amount of time in milliseconds.

Results

Experimenter and Lazy Speaker errors are both miscommunications but imply very different intentions. An *Experimenter Error* indicates the speaker did not realize that the ambiguous statement could have been harmful, unless she received feedback that what she said was wrong or confusing. However, the *Lazy Speaker Error* indicates that the speaker failed to put forth the necessary effort to disambiguate and thus violated the principle of collaborative effort.

The current study aims to answer three questions: 1) How does feedback affect the processing of ambiguous statements over time?; 2) How much effort will a listener put forth to understand a miscommunication?; and 3) Does the amount of effort to understand a previous misunderstanding affect the amount of effort put forth on future language comprehension? We predict that negative feedback should make the ambiguity more salient—thus recruiting more cognitive resources initially (questions 1 & 2). However, as the listener learns that the speaker's ambiguity is often unreliable (causing communication breakdown or failure) and negatively impacts the listener, we should see cognitive effort decline (question 3).

The data was analyzed using growth curve models, multivariate methods for analyzing time series data that simultaneously allows for the measurement of group and individual level effects (Mirman, Dixon, & Magnuson, 2008). The calculation of orthogonal polynomials resolves the issues of dependence associated with the time series. The first orthogonal indexes linear slope, and the second orthogonal indexes line curvature.

Understanding the Ambiguity

A growth curve model evaluated the effects of **Feedback** (*Consequence* vs *No Consequence*) and **Error Type** (*No Error*, *Experimenter Error*, vs. *Lazy Speaker Error*) on dwell time to the "Partner Screen" interest area of the Feedback screen as an indirect measure of how hard the listener tried to understand why the miscommunication occurred (see Fig. 1, *right panel* to reference the "Partner Screen" region of the Feedback Screen). Based on visual inspection of the group data, we decided to calculate up to the second orthogonal polynomial to interact Feedback and Error Type (see Table 1 for statistics and p-values).

The significant main effect of the first orthogonal indicated that as the listeners progressed through the experiment, dwell time significantly decreased—essentially, task adaptation. The main effect of **Error Type** indicated that *Experimenter* and *Lazy Speaker Errors* had longer dwell times than *No Error* trials ($p < .001$) – demonstrating that global ambiguity recruited more cognitive effort.

More interesting is the effect of the global ambiguity over time. The significant **first orthogonal (linear slope) polynomial** x **Error Type** shows a steeper linear decline for the *Experimenter* and *Lazy Speaker Errors* relative to *No Error*. It appears that globally ambiguous statements were seriously considered initially, but over time, the listeners gave up on exerting effort in processing. Potentially, this effect is a result of disengaging from the task. The significant **second orthogonal (line curvature) polynomial** x **Error type** interaction suggests that at least one of the Error Types was marked by a more curvilinear line, which is possibly reflective of adapting to the global ambiguity. Finally, the **Feedback x Error Type** interaction suggests that listeners' dwell times during global ambiguity trials decrease as a function of negative feedback: listeners that were penalized stopped considering the speaker's perspective faster on the global ambiguity trials than listeners that were not penalized (at least $p < .05$; see Fig. 2).

Table 1: Estimates, standard errors, t and p-values for the growth curve model evaluating dwell time to the "Partner Screen" region of the Feedback screen.

Effect	b	se	t
1st orth.	-250.39	45.79	-5.468***
Error	2133.48	223.54	9.54***
1st orth. x Error	-1304.67	206.68	-6.313***
2nd orth. x Error	568.33	172.12	3.30***
Feedback x Error	87.23	18.53	4.47***

Note: ***p < .001

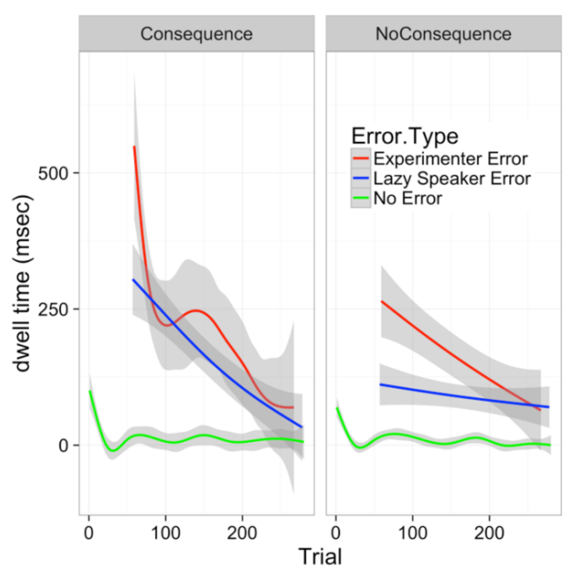

Figure 2: Average dwell time (msec) on the Partner Screen location of the Feedback screen as a function of Feedback Condition by Error Type.

Effect of Global Ambiguity on Future Success

Interestingly, the *No Error* trials had a near-sinusoidal pattern of dwell times (see Fig. 2). The purpose of this growth curve analysis was to determine if the global ambiguity trials had an effect on *No Error* trials. Specifically, we used a growth curve model to explore changes in dwell time during language comprehension (i.e., object locations on the Response Screen, where the listener has to select the target shape). We are especially interested in how hard the listener tried to understand the speaker's perspective after the listener just experienced global ambiguity on the previous trial.

A growth curve model evaluated the effects of **Feedback Condition** (*Consequence* vs *No Consequence*) and the **Error type** (*No Error*, *Experimenter Error*, vs. *Lazy Speaker Error*) on the No Error trials, up to the second orthogonal polynomial. Experimenter Error and *Lazy Speaker Error* trials were excluded from this analysis to reveal the effect of the speaker's intent on listener comprehension of the speaker's future instructions (see Table 2 for statistics and *p*-values).

Table 2: Estimates, standard errors, t and p-values for the growth curve model evaluating dwell time to the object regions of the Instruction screen.

Effect	b	se	t
1st orth.	-2095.03	646.01	-3.24**
Error	207.37	77.28	-2.68**
2nd orth. x Error	-4648.72	1539.35	-3.02**

Note: **p < .01

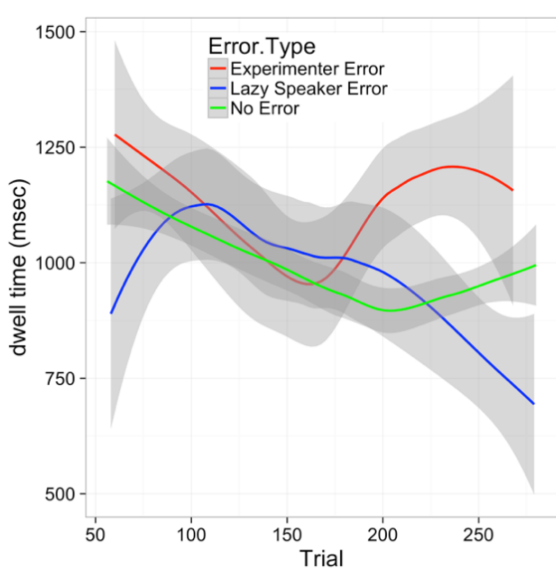

Figure 3: Average dwell time (msec) on the objects on the response screen as a function of Error Type experienced on the previous trial, for No Error Trials.

The main effect of the first orthogonal (linear slope) polynomial demonstrates that dwell time to the speaker's objects on the response screen decreases over time regardless of experimental conditions—i.e., task adaptation. Simple effects analysis of the Error main effect found if a listener had just experienced a *Lazy Speaker Error*, dwell times to the speaker's objects were significantly shorter on the next trial – indicating the *Lazy Speaker Error* negatively affected future language comprehension ($p < .05$). If the listener perceives the speaker lacking in effort, the listener will correspondingly not put in effort to comprehend the speaker's message. However, if the listener perceives the error to not be the fault of the speaker, processing effort increases on the next *No Error* trial ($p < .05$; see Fig. 3). Finally, the interaction between the second (line curvature) orthogonal polynomial and Error type showed that *Lazy Speaker Errors* produced a more curvilinear line than the *Experimenter Error* and *No Error* types ($p < .05$), indicating that a *Lazy Speaker Error* disrupted processing more than the other trial types.

Discussion

In summary, our sample of listeners in typical communication expended different amounts of effort based on the available information. We first evaluated how much effort listeners put forth in considering the speaker's perspective (i.e., "Partner Screen" interest area). Overall, it would seem that when a listener is penalized (i.e., receives negative feedback), the listener more quickly stops considering or reduces consideration for the speaker's perspective. Past errors also modulate future comprehension of the speaker's statements. When a listener experiences a **Lazy Speaker Error** (relative to **Experimenter Error** trials), the listener spends significantly less time considering the objects on subsequent **No Error** trials. It would seem that if a listener understands errors were not the speaker's *fault*, the listener would exhibit willingness to exert more cognitive effort. However, if the speaker is perceived to be *lazy*, listeners will reciprocate – by being *lazy* themselves.

In addition, the very first error presented to participants was an *Experimenter Error*. There were marginally longer processing times on the "Partner Screen" interest area in the *Consequence* condition relative to the *No Consequence* condition ($t = 1.696$, $p = .09$), showing a trend in a direction of longer dwell times when negative feedback is provided. The negative feedback seems to make the listener consider the speaker's perspective, probably because the listener was surprised and wanted to understand what happened. Even more interesting was the drastic decrease in processing time at the last instance of a **Lazy Speaker Error** for the **Consequence** condition ($t = 3.636$, $p < .001$). This suggests that listeners might put forth significantly less effort in trying to understand their partner's perspective if they perceive their partner to not care and were penalized for the error. Therefore, it would seem that listeners and speakers may develop a *tit-for-tat* relationship when negative consequences affect the communicators over the course of an interaction. This is consistent with previous work on listeners handling ambiguous language in a rational manner. For example, Degen, Franke, and Jäger (2013) suggest that listeners interpret ambiguous referents in a game theoretical manner - assuming the ambiguous word refers to only one target because if the speaker meant to refer to the other target, they could have used an unambiguous word instead. When messages had different costs, as unambiguous costs increase, listeners make more inferences based on the ambiguous messages (Rohde et al., 2012). In this case, listeners rationally respond to a speaker that is uncooperative by disengaging from the task, and respond to a speaker that was misinformed with more attention.

Conclusion

Ambiguity need not be problematic for conversations because it can be quite helpful in reducing some of the cognitive effort exerted by both listener and speaker in typical communication. However, these benefits happen only if the ambiguity is properly situated in context. Ambiguity may also promote deeper processing by requiring repair and attention to detail.

It would seem that, in the current study, the interpretation of ambiguity can be ignored when a visual referent helps disambiguate. However, when an ambiguity becomes problematic, the listener weighs the consequence and reason for the ambiguity – subsequently affecting future processing effort. This work provides evidence that listeners are able to interpret miscommunications within the communication context and respond differentially based on both intent and direct consequences.

Acknowledgements

A special thanks goes to the research assistants in the MADI lab at Kent State University.

References

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*(3), 1018–1034.

Degen, J., Franke, M., & Jäger, G. (2013). Cost-based pragmatic inference about referential expressions. *Proceedings of the 35th Annual Conference of the Cognitive Science Society (CogSci'13)*, 376–281.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 641–655.

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

Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. *The Processing and Acquisition of Reference*, *2327*(8957), 239–272.

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*(1), 43–61.

Horton, W. S., & Gerrig, R. J. (2005). The impact of memory demands on audience design during language production. *Cognition*, *96*(2), 127–142.

Jaswal, V. K., & Neely, L. a. (2006). Adults don't always know best. *Psychological Science*, *17*(9), 757–758.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kuhlen, A. K., & Brennan, S. E. (2010). Anticipating distracted addressees: How speakers' expectations and addressees' feedback influence storytelling. *Discourse Processes*, *47*(7), 567–587.

Lane, L. W., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1466–1481.

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Roche, J. M., Dale, R., & Kreuz, R. J. (2010). The resolution of ambiguity during conversation: More than mere mimicry? *Proceedings of the 32nd Annual Meeting of the Cognitive Science Society*, 206–211.

Roche, J. M., Paxton, A., Ibarra, A., & Tanenhaus, M. K. (*under review*). Communication breakdown prompts readjustment of informativeness.

Rohde, H., Seyfarth, S., Clark, B., Jäger, G., & Kaufmann, S. (2012). Communicating with Cost-based Implicature: a Game-Theoretic Approach to Ambiguity. *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue*, 107 – 116.

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes*, *27*, 1–33.