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
Champaign, IL 61801

Zoe Kriegel* (zkriegel@kent.edu)  Jennifer Roche (jroche3@kent.edu)
School of Health Sciences, Kent State University
Kent, OH 44223 USA

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. We usually think of miscommunication exclusively as an outcome to minimize, but what if this view unfairly disparages an element of communication that can be neutral or sometimes even useful? Roche, Paxton, Ibarra, and Tanenhaus (under review) posit that miscommunication can promote deeper processing and more care to resolve misaligned representations of the world. If the presence of miscommunication matters less than the outcome of and reason for the miscommunication, then different types of miscommunication should produce different reactions in a listener.

Ambiguity is woven into the fabric of language. Piantadosi et al. (2012) argued that ambiguity is a natural outgrowth of establishing referents, common ground, and reducing collaborative effort. In some contexts, ambiguous phrases are easily disambiguated by the context, such as saying “look at that!” when seeing something unexpected. 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, it delays comprehension of the listener (Metzing & Brennan, 2003). Ambiguity can be a tool to streamline communication, but only when used effectively with enough context.

A mismatch between perceived context and actual context can result in miscommunication. For example, a first time instructor may have difficulty recognizing that her students may lack foundational information needed to disambiguate higher level concepts (e.g., ANOVA). If the instructor is able to notice the confusion of her students, the miscommunication serves as necessary feedback to adjust the degree of assumed knowledge. However, if the students repeatedly ask questions to confirm understanding and receive contradictory or incorrect answers, the students might disengage and no longer attempt to learn. Relatedly, while preschoolers usually rely upon adults over their peers for accuracy of new information, detected inaccuracies from an adult quickly override this tendency (Jaswal & Neely, 2006). In the study by Birch, Vauthier, and Bloom (2008), even young children show capabilities to learn selectively from more reliable sources without specifically discussing prior reliability of relevant speakers. Speakers that continually provide false or vague information may cause listeners to disregard subsequent utterances by the same speaker.

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. A large body of literature suggests that listeners can flexibly incorporate contextual information to intuit the intentions of speakers.

* Craycraft and Kriegel contributed equally to the preparation of this manuscript and share first author position.
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). Listeners also rely on extra-linguistic cues to make predictions about a speaker’s perspective (e.g., disfluencies produced by a speaker; Heller, Arnold, Klein, and Tanenhaus, 2015). Listeners suspend their expectations of correct usage when a demonstrably unreliable speaker who mislabeling common objects misuses modifiers (Grodner & Sedivy, 2011). Speakers also have assumptions about listeners: namely, that they will pay attention. In Kuhlen & Brennan (2010), speakers told jokes they had previously learned. When speakers expected attentive listeners and received distracted listeners, they used fewer details. We know that speakers are sensitive to listener characteristics, but we do not know to which degree the speaker’s characteristics might influence listener attention and effort. In the current study, we hope to shed light on the listener’s sensitivity to speaker characteristics.

In the following task, listeners 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. The listener may more easily “forgive” speaker ambiguity if it was unintentional, such as when perspectives mismatched and the speaker did not know ahead of time. However, if the listener perceives the speaker to provide ambiguous language because they were being lazy by not putting forth the effort to disambiguate, the listener may expend less effort trying to understand the speaker’s intentions over time. If listeners judge miscommunications in part by the consequences of them, listeners that receive negative feedback will be more likely to disengage from the task than the listeners that do not, especially when the error is from a lazy speaker. We expect listeners to process ambiguity differently depending on the reason for the ambiguity and the effects of ambiguity on communicative success.

**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 (details below). 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. All participants were native speakers of American English with normal to normal-corrected vision. None reported speech or hearing impairments.

**Materials & Stimuli**

A 21 inch iMac (experiment computer), Eyelink 1000 (eye-tracking/host 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 saw 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 (see Table 1 for Feature Type).

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Object Pairing</th>
<th>Feature Overlap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>![Color]</td>
<td>1</td>
</tr>
<tr>
<td>Shape</td>
<td>![Shape]</td>
<td>1</td>
</tr>
<tr>
<td>Size</td>
<td>![Size]</td>
<td>1</td>
</tr>
<tr>
<td>Color &amp; Shape</td>
<td>![Color &amp; Shape]</td>
<td>2</td>
</tr>
<tr>
<td>Color &amp; Size</td>
<td>![Color &amp; Size]</td>
<td>2</td>
</tr>
<tr>
<td>Shape &amp; Size</td>
<td>![Shape &amp; Size]</td>
<td>2</td>
</tr>
<tr>
<td>Color, Shape, &amp; Size</td>
<td>![Color, Shape, &amp; Size]</td>
<td>0</td>
</tr>
</tbody>
</table>

**Auditory stimuli.** Participants were presented with pre-recorded statements from a Caucasian female talker that referenced one, two, or three possible features of the visual stimulus (see Table 2 for example). 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, which did not affect recording quality.
**Table 2: Examples of pre-recorded statements in reference to overlapping elements.**

<table>
<thead>
<tr>
<th>Referent Overlap</th>
<th>Type</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>Shape</td>
<td>&quot;...triangle.&quot;</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>&quot;...big shape.&quot;</td>
</tr>
<tr>
<td></td>
<td>Color</td>
<td>&quot;...red shape.&quot;</td>
</tr>
<tr>
<td>Two</td>
<td>Color &amp; Size</td>
<td>&quot;...big red shape.&quot;</td>
</tr>
<tr>
<td></td>
<td>Color &amp; Shape</td>
<td>&quot;...red triangle.&quot;</td>
</tr>
<tr>
<td></td>
<td>Shape &amp; Size</td>
<td>&quot;...big triangle.&quot;</td>
</tr>
<tr>
<td>Zero</td>
<td>Color, Shape, &amp; Size</td>
<td>&quot;...big red triangle.&quot;</td>
</tr>
</tbody>
</table>

**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 was seated at a computer and microphone around the corner. The listener and confederate received separate instructions, but the listener could hear the experimenter provide the confederate with instructions. The listener and confederate were told they would be interacting 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 saw 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, this 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, *right 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 the 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.

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, listeners were 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. 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 the correct answer.

**Measures**

Over the course of the experimental trial, dwell times were evaluated (i.e., how long the listener fixated within an interest area). The three areas of interest included the two objects on the Instruction Screen and the location of the “Partner Screen” information on the Feedback Screen. Dwell times were calculated using the Eyelink Dataviewer software based on a fixation that landed in the interest area for a predetermined amount of time in milliseconds. 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).
Results

Experimenter and Lazy Speaker errors are both miscommunications but with very different intentions. An Experimenter Error indicates that 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.

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).

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 and negatively impacts the listener, we should see cognitive effort decline (question 3).

The data was analyzed using growth curve models, which are multivariate methods for analyzing time series data that simultaneously allows for the measurement of group and individual level effects (Mirman, Dixon, & Magnuson, 2008). This is achieved through the calculation of orthogonal polynomials, resolving 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 Figure 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 (first orthogonal = indices of linear slope; second orthogonal = curvature of the line).

The results indicated two main effects: first orthogonal (linear slope) polynomial (b = -250.388, se = 45.789, t = -5.468, p < .001) and Error Type (b = 2133.477, se = 223.535, t = -9.544, p < .001) and three interactions: first (linear slope) & second (line curvature) orthogonal polynomial x Error Type (1st orthogonal: b = -1304.668, se = 206.675, t = -6.313, p < .001; 2nd orthogonal: b = 568.331, se = 172.115, t = 3.302, p < .001) and Feedback x Error Type (b = 87.233, se = 19.529, t = 4.467, 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.

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 (b = 125.409, se = 9.765, t = 12.843, 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 No Error trials were 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 decreased 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 Figure 2).

**Effect of Global Ambiguity on Future Success**

Interestingly, the No Error trials had a sinusoidal pattern of dwell times. 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 the speaker’s intent on listener’s comprehension of the speaker’s future instructions.

The results indicated two main effects: first orthogonal (linear slope) polynomial ($b = -2095.03$, $se = 646.01, t = -3.24, p < .01$), Error Type ($b = -2073.37$, $se = 77.28, t = -2.68, p < .01$), and an interaction between the second orthogonal (line curvature) polynomial x Error ($b = -4648.72$, $se = 1539.35, t = -3.02, p < .01$). 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 ($b = -174.23$, $se = 78.63, t = -2.22, p < .05$). If the listener perceives the speaker not putting in effort, the listener will then 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 ($b = -136.72$, $se = 56.77, t = -2.41, p < .05$; see Figure 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 ($b = -3393.82$, $se = 1592.94, t = -2.13, p < .05$), indicating that a Lazy Speaker Error disrupted processing more than the other trial types.

![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.](image)

**Discussion**

In summary, listeners expend 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 considers the speaker’s perspective less. 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 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 over time, listeners that receive negative feedback consistently put forth significantly less effort in trying to understand their partner’s perspective if they perceive their partner to not care, and they 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, Seyfarth, Clark, Jaeger, & Kaufman, 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 to reduce some of the cognitive effort exerted by both listener and speaker. 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, at least 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 will weigh the consequence and reasons for the ambiguity – which will subsequently affect future processing effort. This work provides evidence that listeners are able to interpret miscommunications within the context that causes them to arise and respond to them 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