Learning from uncertainty: exploring and manipulating the role of uncertainty on expression production and interpretation

Amanda Pogue (apogue@ur.rochester.edu) & Michael K. Tanenhaus (mtanenha@ur.rochester.edu)
Department of Brain and Cognitive Sciences, University of Rochester
Rochester, NY 14627 USA

Abstract

Linguistic devices that mark confidence (uncertainty) have been well documented (e.g., choice of modals, hedges, intonational contour), there has been surprisingly little empirical work that explicitly measures how uncertainty is signaled and interpreted. We present an initial report on a project designed to investigate how interlocutors communicate uncertainty and use that information in acquiring new information and integrating interlocutor based input with their prior beliefs. Experiment 1 establishes that speakers and listeners agree on the relative degree of uncertainty for a set of phrases. Experiment 2 manipulated how likely it was that a participant would recognize an object using images that varied in recoverability, finding that recoverability mapped onto certainty. Experiment 3 used a word-learning paradigm in a betting game demonstrating that certainty influenced: the size of a bet that listeners would place on the correctness of a label for a novel object; but not the accuracy of their choice.

Keywords: language; communication; learning; uncertainty; pragmatics

Introduction

Successful communication requires speakers to signal their degree of confidence (certainty) about an utterance through lexical choices and prosodic markings (e.g., It's a dog, I think/THINK it's a dog) and listeners to successfully use these signals and adapt to variations in how different speakers signal uncertainty and whether a speaker is likely to overestimate or underestimate how certain she should be. Understanding how interlocutors mark and interpret confidence in particularly important because a growing body of research in pragmatics demonstrates the information about source, in particular, which speaker is likely to have the more reliable knowledge about information that is being added to common ground, strongly affects the form of utterances (Bibyk, 2016; Gunlogson, 2008).

Linguistic devices that can signal degree of confidence are well-documented. They include choice of modal, adverb, use of hedges, and intonational contour. However, there is a surprising dearth if empirical work that (a) directly relates objective measures of how confident a speaker should be with the signals she sends and (b) measures how the speaker's confidence affects the listener's use of information in the utterances.

We are beginning to fill this gap in the literature using a four-pronged research strategy. First, we scale the relative confidence for a set of expressions. Second, we ground the relative confidence a speaker *should* have in information she conveys by manipulating visual factors known to affect the likelihood that visual/perceptual information is extracted from the input, and determine whether that maps onto these linguistic expressions. Third, we examine how a speaker's use of an uncertainty expression modulates a listener's behavior. Fourth, we examine how interlocutors adapt to how the other weighs and signals their degree of certainty in in goal-oriented communication tasks.

In the current paper we present an initial set of results that establish a proof-of-concept for this research strategy, focusing on the first three prongs. We first establish that listeners have a stable preference for a set of different lexical structures that mark uncertainty (Experiment 1). Next, we introduce a task in which that asks speakers to produce labels for visually displayed objects they are likely to be perceptually uncertain about, in order to see how perceptual uncertainty influences the utterances speakers choose for communicating their labels to another interlocutor (Experiment 2). We then examined how the use of linguistic uncertainty cues maps onto other behaviors, specifically by looking at how they affect learning names for novel objects using a word-learning betting paradigm adapted from research on decision-making (Experiment 3). Together these studies establish a methodological and empirical foundation for future work that systematically examines how interlocutors convey uncertainty.

Experiment 1: Pre-testing materials

This experiment was conducted to test whether we could devise a set of phrases that a speaker would use to mark differences in certainty. We tested eight phrases that we thought marked different levels of un/certainty. We also examined whether listeners would be able to reconstruct the certainty of a speaker who used the phrases.

We tasked several naive speakers with producing these sentences: by first reading them out loud, and subsequently reading them out loud while imagining that they might be uncertain of the correct label for the relevant item. We predicted that speakers would modulate their speech when they were asked to mark uncertainty compared to just reading aloud the same sentences. A different set of participants were then asked to rate how certain the speaker was, by listening to these recordings, or by reading each of the phrases.

Methods

Participants 8 naive graduate students and research staff from the authors' department were recruited to record sentences. Participants knew that their recordings were being used for a study, but were not informed of the purpose or hypotheses being tested. Participants were all native

speakers of American English. An additional 176 participants were recruited using Amazon's Mechanical Turk (MTurk), and were compensated $0.25 for completing the task. All participants were self-reported native speakers of English.

Recording Procedure Participants were seated at a desk in front of a laptop. They were told that they would see several sentences appear one at a time in a random order, and that their job was to read each sentence out loud. They could then click the "next" button to proceed to the following sentence. After recording all eight sentences (see: Table 1), participants were asked to read the sentences again, also in random order, but this time imagining that it was difficult to identify the object in the sentence, either because it was partially occluded, or had been quickly presented. Thus, that they might be uncertain about whether the label they used was correct. Recordings were made using the built-in speakers on a MacBook Pro. Each sound file was trimmed into individual phrases.

Rating Procedure Participants were told they would hear (Listen Condition) or read (Read Conditions) sentences from previous Turkers (workers on MTurk) who were asked to identify pictures of birds, that may have flashed quickly or may have been degraded in some way. Their task was to rate how confident the speaker sounded on a 100-point scale, where 0 is not at all confident and 100 is completely confident. Each person heard or read one instance of each of the phrases. The order of phrases, which speaker produced that phrase, and whether the phrase was in read-speech or uncertainty-speech was randomized. After completing all of the ratings participants were asked to rank the phrases (written) in order of certainty.

Results and Discussion

Regardless of condition we find a similar order between the rated certainty, and relative rankings (results listed in Table 1). This demonstrates that: (a) listeners can mark differing amounts of uncertainty with their lexical choice, and (b) the phrases we created mark different levels of certainty.

Experiment 2: mapping visual certainty onto linguistic certainty

To investigate whether speakers similarly pick phrases to express their un/certainty, we directly manipulated the perceptual certainty a speaker should have by manipulating the completeness of the images presented to her and the duration of time she had to view that image. This permitted us to test how that speaker would naturalistically mark their certainty. We used images from a classic perceptual recognition study (Biederman, 1987), that are known to be more-or-less difficult to identify at short exposure durations (intact/recoverable/non-recoverable line drawings). We presented the images to participants at several varying short exposure durations. We expected that both the duration of presentation and the recoverability of the image would affect both how accurately the speaker identified the pictured object, and the speaker's certainty of in their label. We also predicted a relationship between that certainty and the likely phrase that the speaker would use to communicate what she saw, and we expected that relationship to reflect the listener ratings in Experiment 1.

Methods

Participants. 145 participants were recruited using MTurk, and were compensated $0.50. All participants were self-reported native speakers of English.

Procedure. Participants were told they would see images appear quickly on the screen, and their task was to identify the image. After viewing the image they were asked to label what they saw. Half of the participants were then asked to rate their confidence in that label. After submitting this information, their label was piped into the 8 phrases from the Experiment 1, and they were asked to select which of the phrases they would chose to describe what they saw to another person.

Each participant saw five items: a mug, a glass, a watering can, a pair of scissors, and a stool. There were three renditions of each image: fully intact, recoverable, and non-recoverable (according to the original study at short

Table 1: Phrases used, confidence ratings, and rank orderings across experiments

Phrase	Exp1 Read-text Confidence	Exp 1 Listen (Read)	Exp 1 Listen (Uncertainty)	Exp 1 Read Rank	Exp 1 Listen Rank	Exp 2 (mean confidence)	Exp 3 (mean confidence)
1. It could be a goose	36.994	37.706	36.283	7.125	7.063	25.163	24.68
2. It might be a robin	39.294	41.094	37.494	6.375	6.375	28.798	
3. I think it's a falcon	49.918	48.918	50.919	5.688	5.644	46.458	25.46
4. It looks like a hummingbird	57.080	61.362	52.797	5.25	5.381	45.828	
5. I'm pretty sure it's a woodpecker	65.476	68.110	62.842	4.063	4.319	68.577	
6. I'm sure that it's a sparrow	84.220	87.510	80.930	2.688	2.919	80.300	
7. It's a blackbird	86.777	88.864	84.689	2.625	2.525	91.765	80.40
8. It's definitely a canary	90.935	90.246	91.624	2.188	1.775	93.192	
Control							64.55

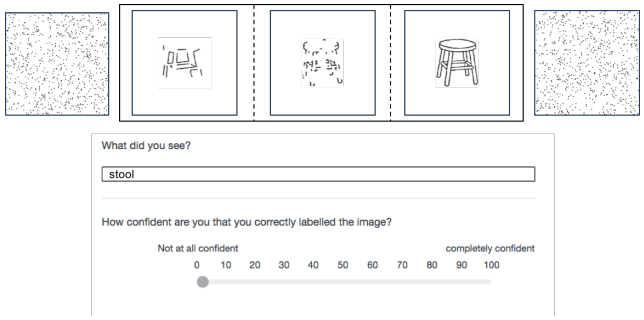

Figure 1: Example display stimuli from Experiment 1 with a non-recoverable, recoverable, and intact picture of a stool (adapted from Biederman, 1987)

durations the line deletions in the non-recoverable images made the image more difficult for participants to identify, than deletions for the recoverable images; see: Figure 1). The images were presented for either 120, 220 or 750 ms between two random dot arrays to prevent participants from recreating the image from the afterimage of the display. Each participant saw only 5 trials (one trial per possible image). The order of presentation, the recoverability of each item, and duration of presentation were completely randomized for each participant.

Results

We find evidence to support our manipulations of certainty (accuracy of label, and mention of visual uncertainty; Figure 2). We see an increase in correct labels by participants when the image is recoverable and they view the images for longer durations. Participants report less visual uncertainty (e.g., reporting that they saw some dots, random lines, something unrecoverable, etc.) for recoverable versus non-recoverable items, and at longer rather than shorter viewing times.

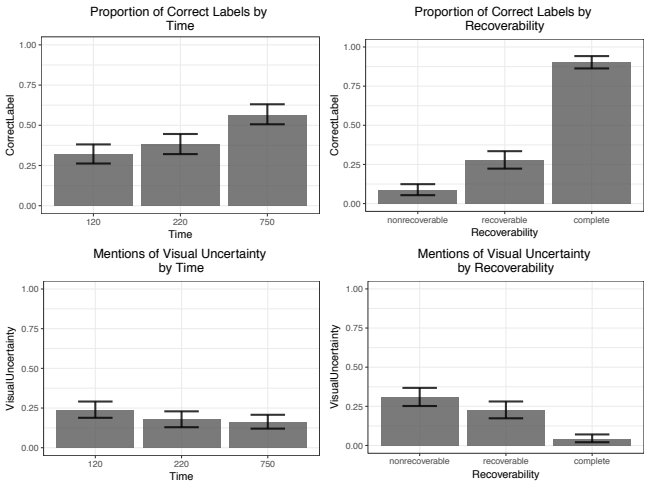

Figure 2: Effects of duration and recoverability manipulations on correct labels and uncertainty

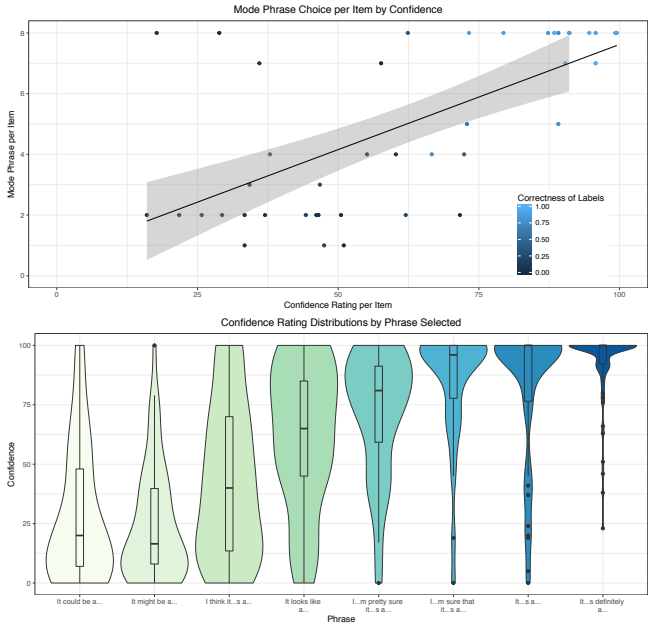

Figure 3: Relationship between phrase chosen by half of the participants and the certainty ratings from the other half by item (top), and the mapping of confidence onto phrase selected (bottom)

We also find a relationship between the certainty judgments, and the phrases chosen by the other half of participants (Figure 3).

Discussion

Taken together the findings of Experiment 1 and Experiment 2 suggest that speakers choose their utterances in a systematic way to express varying amounts of uncertainty. In addition, listeners are able to accurately assess that uncertainty. Importantly, speakers can express their perceptual uncertainty with linguistic uncertainty. This provides us with a foundation for using visual stimuli (e.g., moving dot patterns) where it is possible to quantify perceptual uncertainty. It also allows us to ask how linguistic uncertainty maps onto behavior in other domains.

Experiment 3: Effects of uncertainty on word learning and memory

Thus far we have seen that speakers and listeners agree on the relative certainty of a set of phrases, but we have yet to explore how uncertainty cues affect behavior in communicative settings. In the current experiment, we ask participants to evaluate a speaker's knowledge state about the correct label for an object. We extend beyond the current work in the field suggesting that listeners are sensitive to cues to speaker knowledge (see: Brennan & Williams, 1995; Krahmer & Swerts, 2005; Smith & Clark, 1993; Swerts & Krahmer, 2005; Swerts, Krahmer, Barkhuysen, & van de Laar, 2003), to show that this allows listeners to adjust their expectations about an interlocutors' referential knowledge.

We devised a task based on findings in the memory, metacognition, and decision-making literature. Prior work suggests a number of factors influence the likelihood that an individual will admit that they know some things with more confidence than others, despite their accuracy on both sets of things being equal (Busey & Tunnicliff, 2000). Several studies demonstrate that perceptual properties such as clarity, font size, luminance, and even the volume of information at encoding influence the likelihood that a participant will recall learning some information (Koriat, 2007; Rhodes & Castel, 2009). We extended this approach to word learning, predicting that listeners will judge their own knowledge about a word-object pairing based on the amount of linguistic certainty expressed by the speaker during learning. We expected that listeners would be more certain of a word-object pairing when the label was presented in an utterance that signals higher certainty.

In order to focus on the listeners' implicit sense of their own knowledge, we borrow a method from the decision-making literature, (Budescu, Weinberg, & Wallsten, 1988; Heath & Tversky, 1991), asking participants to bet on their object-label pairing judgments. If a listener is certain that they have correctly accepted or rejected an object-label pair they should place a maximum bet. As certainty decreases, they should adjust their bet to maximize their likely reward, and minimize their losses.

Methods

Participants 75 participants were recruited using MTurk, and were compensated $0.35 for completing the norming task. An additional 64 participants were recruited using MTurk. They were compensated $0.60 with a bonus of up to $0.60 for completing the word-learning task. All participants were self-reported native speakers of English.

Materials The 12 novel objects used in the learning task were taken from 3 distinct families of Fribbles (Williams, 1998). The names for each of the Fribbles were recorded using phrases 1, 3, and 7 ("That could be a…", "I think that's a …", and "That's a…") from the Experiment 1. Bare-noun instructions were created by splicing the noun from the most certain (#7) recordings.

Norming Procedure: Participants were told that previous Turkers were taught the names of 16 novel objects. They were told that at various stages of learning, these Turkers had been asked to produce the label for each of the objects. In the current task, participants were told that they would be shown the object that the speaker was asked to label, and would hear the label that was produced by that speaker. Their task was to evaluate how certain the speaker sounded about their label on a scale of 1 to 100, where 1 was not at all certain, and 100 was completely certain. Participants heard the labels for 12 objects, in pseudo-randomized orders. Each participant heard 4 objects labeled with each of the 3 expressions. A quarter of the participants heard just the bare-noun expressions. Participants only heard one labeling event for each object, and were only able to play the audio label once. Expression-object pairing was counter-balanced across conditions.

Word-learning Procedure: Participants were taught the names of the 12 distinct Fribbles, using the same recordings as in the norming task. They were told that the names were recorded by a previous Turker, who learned the names of 16 objects. As in the norming task they were told that they would hear the labels from different stages of learning. Participants were told that they would hear a label for 12 of the 16 items, and that they would only hear the label once. After hearing each of the labels they would be asked to bet on whether or not the correct label was used.

Participants saw each item separately and could click on a button to hear the recorded label. They were instructed that they would not be able to hear the label more than once (the play button turned gray and could not replay the sound after the first button push). Each participant heard 4 objects labeled using each of the 3 expressions. A quarter of the participants heard just the bare-noun. Items were presented in a pseudo-randomized order and expression-object pairing was counter-balanced across conditions.

During the test phase, participants were shown each of the items one at a time, with a written label. They were asked first if they thought the label was correct, and then they were asked to bet up to $0.05 that their guess was correct (e.g., that the label was in/correct). If they were correct their bonus would increase by the amount they had bet, but if they were incorrect the amount they had bet would be taken out of their bonus pot. They were instructed that this money was a bonus, and that their account would be credited for the amount in their bonus pot after they completed the task, but that they would not lose their base pay if their bonus was below zero. We predicted that participants would bet more money for items they felt the most confident about (either being correct or incorrect). Half of the object-label pairings were correct.

Results

Norming Results: The mean certainty ratings for each of the expressions are presented in Table 1. Overall we find that listeners thought that the speaker sounded the most certain for statements such as "That's an X", followed by the bare-noun (control) statements. The more uncertain phrases (1 & 3) were rated similarly uncertain, and their combined rating was less certain than the other two phrases ($p < .001$). As predicted by the previous experiments we find a relationship between the phrase used and the speaker's perceived certainty in their label.

Word-learning Results Listeners placed higher bets for the items that were described with more confidence ($ps < .001$). However, confidence did not make onto accuracy, mirroring other results from the memory literature. ($ps > .1$; Figure 4).

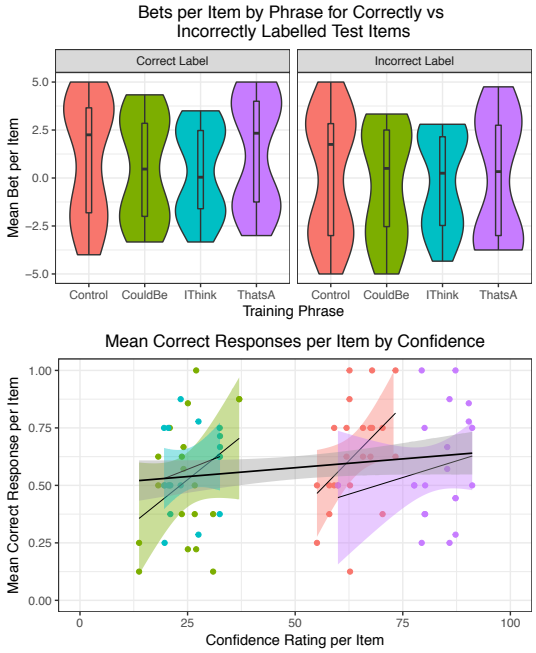

Figure 4: Betting data by participants in Experiment 3 by Training Phrase type for Test items, and Accuracy by Confidence (from Experiment 2)

Discussion

The data from this experiment suggest that while the fluency of the signal may not improve learning accuracy (as predicted by the metacognition and memory literature), listeners are making use of the certainty information. The certainty information provided by the speaker influences the extent to which a listener feels confident about an object-label pairing, suggesting that they are more willing to accept a confidently labeled item as being the correct label, than a label marked with uncertainty.

General Discussion

The evidence from this series of studies is a first step in developing methodologies to evaluate how speakers convey uncertainty and how listeners modulate their expectations based on the inferred uncertainty of a speaker in a dynamic communicative setting. We provided several proof-of-concept studies to show that speakers can make use of different cues to uncertainty, and that listeners can reliably interpret that uncertainty (Experiment 1 and 2). We have shown that we can employ methodologies that allow us to map speakers' linguistic uncertainty onto their perceptual uncertainty. We also show that we can extend this to applications from other fields such a psychophysics, memory, decision-making, and word-learning to examine how linguistic certainty affects behaviors in these domains. This is demonstrated in Experiment 3 where borrowing methods from other cognitive tasks allowed us to investigate the role of linguistic certainty on a listener's judgment of their own knowledge, and their memory representations. This work establishes methods that allow us to directly manipulate and test both speakers' and listeners' certainty and also manipulate and test the behavioral effects of that certainty. This work sets the stage for future work that investigates the relationship between certainty marked in an utterance, and the kinds of inferences that are drawn by both the speaker and the listener. We conclude by briefly describing future work, some of which is ongoing.

Current and Future Directions

Conflicting evidence and uncertainty In a study in progress we aim to replicate some of the work in the classic word learning literature (Koenig & Harris, 2005; Scofield & Behrend, 2008; Vanderbilt, Heyman, & Liu, 2014), pitting a more certain speaker against less certain speaker. In this line of research we aim to investigate how listeners are able to use certainty to consider the reliability of a source. We also attempt to examine how this information can cue expertise, and how this affects perceived reliability.

Evaluating speaker knowledge In a current study, following up on previous work (Ibarra, Runner & Tanenhaus, 2017) about co-operative communication, we manipulate the perceived expertise or uncertainty of a speaker, and evaluate how an interlocutor will modulate their future expressions when talking about a given topic with that speaker. For example, if a speaker believes that their interlocutor has expert knowledge about kitchen utensils they might be more willing to use the proper names for utensils they have privileged knowledge about, indicating that they have made a generalization about their interlocutor's likely knowledge. By comparison if their interlocutor shows more uncertainty about the labels of the same items, a speaker might infer that they have less knowledge in that domain, and might instead choose to describe rather than name an item they have privileged knowledge about (as seen in studies on shared vs. privileged ground in reference; see: Gegg-Harrison & Tanenhaus, 2016; Gorman, Gegg-Harrison, Marsh, & Tanenhaus, 2013; Heller, Gorman, & Tanenhaus, 2012).

Reliability of uncertainty cues Our previous work has shown that interlocutors have expectations for how people will typically refer to things in the world, and that they can flexible adapt these expectations speaker-specifically for speakers that deviate from the norm (Pogue, Kurumada, & Tanenhaus, 2016). These set of tasks combine these two lines of research to ask how interlocutors determine whether a speaker is deviating from an expected use of certainty cues (e.g., a speaker may mark uncertainty, despite being fully knowledgeable, or, conversely may mark certainty for things they have little knowledge on; see: nervousness vs. mansplaining), and how they might adapt to these deviations. This line of research also asks how speakers flexibly adapt to listeners who are mis/interpreting their certainty cues.

Summary

The studies discussed in this paper demonstrate that we can manipulate the degree of certainty an interlocutor has for a given piece of information. We can then measure how speakers linguistically communicate this certainty, and how listeners make use of that information. The work provides a foundation for future work on how interlocutors mark certainty in their interactions with each other, and how they might update their utterances and expectations by taking into account how, and with what degree of reliability, their interlocutor signals or interprets uncertainty.

Acknowledgments

The authors would like to acknowledge the members of the Human Language Processing (HLP) Lab, and attendees of XPrag 2017 for their helpful feedback on these projects. AP was funded by a fellowship through the University of Rochester Center for Language Sciences for a portion of the duration of these studies.

References

Bibyk, S. A. (2016). *A Rise by Any Other Name: An Investigation of the Production and Comprehension of Rising (and Falling) Intonation in Questions*. *Dissertation*.

Biederman, I. (1987). Recognition-by-Components : A Theory of Human Image Understanding Theoretical Domain : Primal Access to Contour-Based Count Versus Mass Noun Entities : The Role. *Psychological Review*, *M*(2), 115–147.

Brennan, S. E., & Williams, M. (1995). The Feeling of Another′s Knowing: Prosody and Filled Pauses as Cues to Listeners about the Metacognitive States of Speakers. *Journal of Memory and Language*. http://doi.org/10.1006/jmla.1995.1017

Budescu, D. V., Weinberg, S., & Wallsten, T. S. (1988). Decisions Based on Numerically and Verbally Expressed Uncertainties. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 281–294. http://doi.org/10.1037/0096-1523.14.2.281

Busey, T. A., & Tunnicliff, J. (2000). Accounts of the confidence – accuracy relation in recognition memory. *Most*, *7*(1), 26–48.

Gegg-Harrison, W. M., & Tanenhaus, M. K. (2016). What's in a Name? Interlocutors Dynamically Update Expectations about Shared Names. *Frontiers in Psychology*, *7*, 1–22. http://doi.org/10.3389/fpsyg.2016.00212

Gorman, K. S., Gegg-Harrison, W., Marsh, C. R., & Tanenhaus, M. K. (2013). What's learned together stays together: Speakers' choice of referring expression reflects shared experience. *Journal of Experimental Psychology: Learning Memory and Cognition*, *39*(3), 843–853. http://doi.org/10.1037/a0029467

Gunlogson, C. (2008). A question of commitment. *Belgian Journal of Linguistics*, *22*, 101–136. http://doi.org/10.1075/bjl.22.06gun

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4*(1), 5–28. http://doi.org/10.1007/BF00057884

Heller, D., Gorman, K. S., & Tanenhaus, M. K. (2012). To name or to describe: shared knowledge affects referential form. *Topics in Cognitive …*, *4*(2), 290–305. http://doi.org/10.1111/j.1756-8765.2012.01182.x

Ibarra, A., Runner, J.T., & Tanenhaus, M.K. (2017). The role of prior knowledge and expertise on choice of referring expression. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. http://doi.org/10.1111/j.1467-8624.2005.00849.x

Koriat, A. (2007). Metacognition and Consciousness Metacognition and Consciousness Metacognition and Consciousness. *Cambridge Handbook of Consciousness*. http://doi.org/http://dx.doi.org/10.1017/CBO9780511816789.012

Krahmer, E., & Swerts, M. (2005). How Children and Adults Produce and Perceive Uncertainty in Audiovisual Speech. *Language and Speech*, *48*(1), 29–53. http://doi.org/10.1177/00238309050480010201

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-Specific Generalization of Pragmatic Inferences based on Under- and Over-Informative Prenominal Adjective Use. *Frontiers in Psychology*, *6*(January), 1–18. http://doi.org/10.3389/fpsyg.2015.02035

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*(3), 550–554. http://doi.org/10.3758/PBR.16.3.550

Scofield, J., & Behrend, D. a. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*(2), 278–290. http://doi.org/10.1016/j.cogdev.2008.01.003

Smith, V. L., & Clark, H. H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, *32*(1), 25–38. http://doi.org/10.1006/JMLA.1993.1002

Swerts, M., & Krahmer, E. (2005). Audiovisual prosody and feeling of knowing. *Journal of Memory and Language*, *53*(1), 81–94. http://doi.org/10.1016/j.jml.2005.02.003

Swerts, M., Krahmer, E., Barkhuysen, P., & van de Laar, L. (2003). Audio visual cues to uncertainty. *Science*, 25–30.

Vanderbilt, K. E., Heyman, G. D., & Liu, D. (2014). In the absence of conflicting testimony young children trust inaccurate informants. *Developmental Science*, *3*, 443–451. http://doi.org/10.1111/desc.12134

Williams, P. (1998). Representational organization of multiple examplars of object categories, 1–21.