# Uncertainty can increase explanatory credibility

Sangeet Khemlani[1*], Daniel Gartenberg[2], Kun Hee Park[2], and J. Gregory Trafton[1]

[*]National Research Council Postdoctoral Research Associate
[1]US Naval Research Laboratory, Washington, DC 20375 USA
[2]George Mason University, Fairfax, VA 22030 USA

## Abstract

In daily conversations, what information do people use to assess their conversational partner's explanations? We explore how a metacognitive cue, in particular the partner's confidence or uncertainty, can modulate the credibility of an explanation. Two experiments showed that explanations are accepted more often when delivered by an uncertain conversational partner. Participants in Experiment 1 demonstrated the general effect by interacting with a pseudo-autonomous robotic confederate. Experiment 2 used the same methodology to show that the effect was applicable to explanatory reasoning and not other sorts of inferences. Results are consistent with an account in which reasoners use relative confidence as a metacognitive cue to infer their conversational partner's depth of processing.

**Keywords:** explanations, confidence, uncertainty, collaborative reasoning, human-robot interaction

## Introduction

What makes an explanation believable? Researchers have recently discovered several conceptual and structural properties that distinguish credible explanations (for reviews, see Keil, 2006; Lombrozo, 2006). Good explanations are often relevant and informative (Grice, 1975; Wilson & Sperber, 2004). Likewise, people appear to prefer explanations that are simple (Chater, 1996; Lagnado, 1994; Lombrozo, 2007; but cf. Johnson-Laird, Girotto, & Legrenzi, 2004), and in situations of uncertainty, they appear to prefer explanations that have narrow *latent scope*, i.e., those that account for only observed phenomena (Khemlani, Sussman, & Oppenheimer, 2011). These preferences show that properties intrinsic to the explanation itself can cause individuals to judge the explanation to be better, more likely, more plausible, and more credible.

However, individuals also rate explanations by appealing to extrinsic information, e.g., information about the context in which the explanation was provided rather than the material content described by the explanation. Extrinsic information is particularly important when reasoners have to evaluate another individual's explanations. In those situations, factors such as the individual's motivation, mood, and confidence can affect the believability of his or her explanation. In this paper, we focus on how confidence can modulate an explanation's credibility. We first describe confidence as a metacognitive signal, and then explain how confidence can affect the believability of an explanation. Two studies show that when an agent appears uncertain, individuals accept the agent's explanations more often. We discuss the phenomenon in light of intuitive and analytic reasoning systems.

## Confidence and explanatory credibility

Subjective confidence is among the most widely investigated metacognitive signals (Dunlosky & Metcalfe, 2009). In many cognitive tasks it is correlated with accuracy, though people are often systematically overconfident about their performance (Lichtenstein, Fischhoff, & Phillips, 1982; Lindley, 1982; McClelland & Bolger, 1994). Much of the research on subjective confidence addresses how individuals integrate cues from their task performance or else their declarative knowledge to assess their confidence in a particular decision of theirs. Confidence is often construed as a signal predictive of translating judgments to actions (Dunning, 2007; Tversky & Koehler, 1994), and researchers have accordingly proposed many models of how that signal is constructed (Albert & Sponsler, 1989; Erev, Wallsten, & Budescu, 1994; Ferrell & McGoey, 1980; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Griffin & Tversky, 1992; Juslin, 1994; Koriat, 2012; May, 1986; Pfeifer, 1994; Wallsten & Gonzáles-Vallejo).

In daily interactions with others, people frequently provide cues to their own level of confidence for their conversational partners to interpret, and they use their partner's cues to interpret the content of their partner's statements. Despite the prevalent use of confidence signals in modulating informational content, little work has established how individuals integrate cues to a partner's confidence or lack thereof into their own decision-making, and few if any of the aforementioned models of subjective confidence can explain how confidence is assessed in others. Suppose, for example, that you ask a friend what she thinks of a new restaurant that has opened up in her neighborhood. If she says, "It's good!" her intonation may provide a cue to a high level of confidence in her response. Alternatively, if she hesitates and says, "It's…good…" then you may negate the material content of her response and prefer instead to explain her lack of confidence as indicative of her disapproval.

In the present investigation, we examined how individuals incorporate their partners' levels of confidence when they assess their partner's explanations of a confusing scenario. Reasoners could modulate their acceptance in their partner's explanation in one of two ways. An intuitive prediction is that people should accept an explanation more often when the explanation is delivered by a confident partner than an uncertain partner. People who exhibit this behavior should infer, implicitly or explicitly, that the partner's confidence is proportional to the explanation's credibility. Preliminary support for this prediction comes from recent studies on so called "powerless language", which show that statements

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| **AUG 2013** | | **00-00-2013 to 00-00-2013** |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT NUMBER |
|---|---|
| **Uncertainty Can Increase Explanatory Credibility** | 5b. GRANT NUMBER |
| | 5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | 5d. PROJECT NUMBER |
| | 5e. TASK NUMBER |
| | 5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory ,Washington,DC,20375** | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. SPONSOR/MONITOR'S ACRONYM(S) |
| | 11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES
**Proceedings of the 35th Annual Conference of the Cognitive Science Society. Cognitive Science Society, Austin, Texas, pp. 2692-2697, 31 Jul ??? 3 Aug 2013.**

14. ABSTRACT
**In daily conversations, what information do people use to assess their conversational partner???s explanations? We explore how a metacognitive cue, in particular the partner???s confidence or uncertainty, can modulate the credibility of an explanation. Two experiments showed that explanations are accepted more often when delivered by an uncertain conversational partner. Participants in Experiment 1 demonstrated the general effect by interacting with a pseudoautonomous robotic confederate. Experiment 2 used the same methodology to show that the effect was applicable to explanatory reasoning and not other sorts of inferences. Results are consistent with an account in which reasoners use relative confidence as a metacognitive cue to infer their conversational partner???s depth of processing.**

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a REPORT **unclassified** | b ABSTRACT **unclassified** | c THIS PAGE **unclassified** | **Same as Report (SAR)** | **7** | |

that include hedging phrases such as "sort of", "kind of", and "probably" are rated more negatively compared to non-hedged statements (Blankenship & Holtgraves, 2005; Durik, Britt, Reynolds, & Storey, 2008; Liu & Fox Tree, 2012). Hedges may provide a cue to a low level of confidence, and therefore cause people to attenuate their belief in the statement.

Alternatively, if people prefer explanations when they are delivered by an uncertain partner, then it may be because the partner's uncertainty provides pragmatic cues to the strength of the explanation. For example, an uncertain expressional cue such as a furrowed brow may suggest that the partner was engaged in more analytical thinking (Alter, Oppenheimer, Epley, & Eyre, 2007), and an analytical response may be preferred to an intuitive one.

In what follows, we report two experiments that tested whether confidence or uncertainty affects explanatory credibility. In both studies, participants engaged in a dyadic interaction with a pseudo-autonomous humanoid robot. The robot allowed us to impose stringent controls on the verbal and expressional cues that participants received.

## Experiment 1

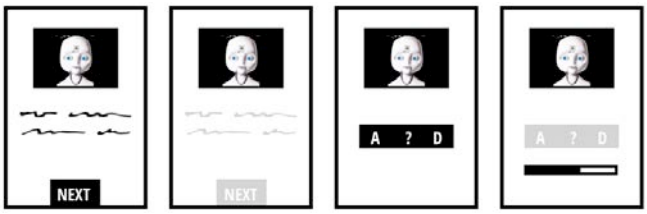
Experiment 1 tested whether an explanation was more or less acceptable if it came from a confident or an uncertain confederate. To generate systematic social interactions, the experiment called on participants to engage in a dyadic interaction with a pseudo-autonomous robotic confederate, a humanoid mobile, dexterous, social (MDS) robot (Breazeal et al., 2008). Participants were told that they were interacting with the robot through a web-based chat interface (see Figures 1 and 2). Participants' task was to read a problem to the robot, listen to the robot's response, and then decide whether they agreed, did not understand, or disagreed with the robot. If they did not understand, or else if they disagreed with the robot, they verbally explained their reason for not accepting the robot's response, and their verbal protocols were recorded. All of the robot's responses were pre-recorded, and we manipulated whether the robot delivered its responses using cues of confidence or uncertainty.

### Method

*Participants.* 38 native-English speaking undergraduates from George Mason University participated in exchange for partial course credit. None of the participants had received any training in logic.

*Procedure.* Participants engaged in a dyadic interaction with a pseudo-autonomous robotic confederate. Before they began the study, they were shown a video of humans engaged in natural language dialogue with an MDS robot (Hiatt et al., 2011). Participants were told that they would interact with the robotic confederate online, but that the confederate had only limited abilities to comprehend natural language, and that the confederate would be unable to respond to unrelated questions. In actuality, all of the robot's responses were pre-recorded. Participants were instructed to use a chat interface to read problems to the confederate and listen to the confederate's responses. The interface was written in Objective C for an iPad tablet computer.

The experiment began when the confederate introduced itself as "Lucas", an MDS robot, and waited for the participant to initiate the study by reading the first problem. Figure 1 shows a schematic of the interface. Participants first read a description of a problem to the confederate (Figure 1a); when they finished, they pressed a button and listened to the confederate's response (Figure 1b); when the robot finished speaking, the participants indicated whether they agreed with, did not understand, or disagreed with the robot's response (Figure 1c); finally, if they disagreed or did not understand the robot, they were given an opportunity to explain their disagreement verbally (Figure 1d), and they moved on to the next problem.




a. b. c. d.

**Figure 1.** A schematic diagram of the chat interface used for the pseudo-interaction in Experiments 1 and 2.

*Design and materials.* Problems consisted of a conditional generalization (1), a categorical statement (2), and an inferential prompt, e.g.,

1. If James does regular aerobic exercises then he strengthens his heart.
2. But, James did not strengthen his heart.
3. What, if anything, follows?

The problems invite both explanatory (e.g., "James had a congenital heart defect") and deductive (e.g., "James did not do regular exercises") responses. However, people tend to elicit explanations for such problems (Lee & Johnson-Laird, 2006). In the present study, participants listened to and evaluated the confederate's explanation of ten separate problems, which were drawn from five different domains: biology, economics, mechanics, psychology, and natural phenomena (see the Appendix for the full set of materials). Explanations were adapted from reasoners' most frequently generated spontaneous explanations in studies that used similar materials (Khemlani & Johnson-Laird, 2012). For each explanation, the robotic confederate delivered its response using a verbal cue and an expressional cue to its level of confidence. Half of the participants received *confident* verbal and expressional cues, and the remaining received *uncertain* cues. The explanations in both conditions were delivered with the same intonation. Figure 2

provides examples of the verbal and expressional cues. The materials were balanced for their length across both conditions.

*Post-experimental questionnaire.* Participants who perceive their interaction with the robot as staged may respond differently than those who believe the interaction is real. To examine factor, participants completed a post-experimental questionnaire after they finished the experiment proper. The questionnaire assessed whether the participants had believed (erroneously) that they were interacting with an autonomous robot, or whether they believed (accurately) that the interaction was staged. In our analyses, we present data from the most direct question they answered, which was as follows:

*"Did Lucas's responses seem natural?*

1. *No, his responses usually looked like pre-recorded videos.*
2. *I'm not sure.*
3. *Yes, he usually responded like a human would."*

After participants answered the questionnaire, they were debriefed that the interaction was staged.

## Results and discussion

Figure 3 shows the percentage of agreement for the explanations as a function of the confederate's confidence. Surprisingly, participants accepted explanations more often when the confederate was uncertain (75% agreement) than when it was confident (63% agreement; one-tailed Mann-Whitney test, $z = 1.75$, $p = .04$, Cliff's $\delta = .33$). In both conditions, participants accepted explanations significantly more often than chance (Wilcoxon tests, $zs > 2.25$, $ps < .02$). Their agreement varied across the different types of materials (Friedman analysis of variance, $\chi^2 = 49.9$, $p < .0001$). Across the study, 45% of the participants responded that they believed the interaction was pre-recorded.

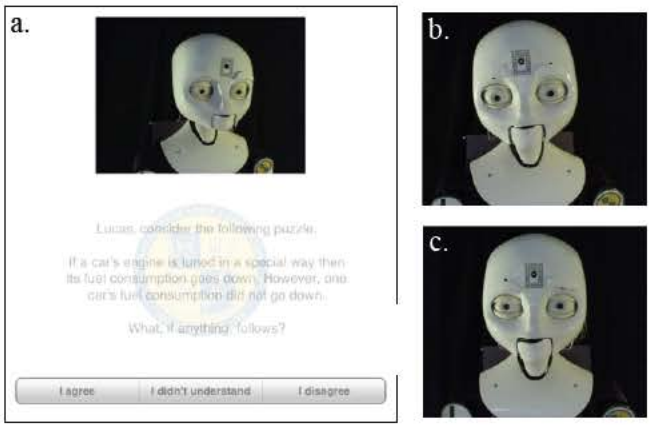



**Figure 2.** The interface used in Experiments 1 and 2 (a). The robotic confederate was either confident (b) or uncertain (c) for the duration of the study. Confident expressional cues included wide open eyes, raised eyebrows, and a straight mouth orientation. Furthermore, the confident confederate preceded its responses with confident verbal cues, e.g., "Oh, I've got it!" or "That's easy." Uncertain expressional cues included narrow eyes, half-cocked eyebrows (a furrowed brow analog), and a slanted mouth orientation. Uncertain verbal cues included expressions such as, "Hmm, that's a tough one" and "Huh, I don't know for sure."

To assess whether the effect of uncertainty on explanatory credibility was robust across the different materials, we fit the data to a generalized mixed-effects model (Baayen, Davidson, & Bates, 2008) with a binomial error distribution and a logit link function using the *lme4* package (Bates, Maechler, & Bolker, 2012) in R (R Core Team, 2012). The model took into account a single fixed effect, i.e., the confederate's confidence, as well as three additional random effects: the participant variance, the problem variance, and whether or not the participant believed that the interaction was pre-recorded. The model yielded a significant main effect of confidence ($b = .77$, $SE = .37$, $p = .04$). The results suggest that the effect held whether or not the participants believed that the interaction was staged.

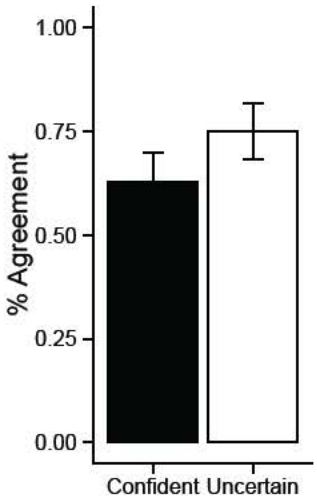



**Figure 3.** Agreement percentages for explanations as a function of whether those explanations were delivered by a confident or an uncertain confederate. 95% confidence intervals shown.

Experiment 1 tested whether reasoners would accept explanations more or less often when given by an uncertain confederate compared to a confident confederate. However, the study did not establish whether the effect is unique to explanatory reasoning. It may be the case that the effect is widespread, and that it is applicable to any sort of inference, not just to the evaluation of explanations. To test the boundary conditions of the effect, participants in Experiment 2 evaluated both explanations and deductions.

## Experiment 2

Experiment 2 sought to replicate the effect of uncertainty on explanatory credibility, as well as to test whether it applied to any sort of inference, or whether it was localized, in part, to explanatory reasoning. The study was similar to the previous one, with one exception: the robotic confederate in the present study provided two types of responses, either an explanation or else a deduction. Recall that the problems used in the previous study, e.g.,

If James does regular aerobic exercises then he strengthens his heart.
But, James did not strengthen his heart.
What, if anything, follows?

invite two different sorts of reasoning strategies. One could construct an explanation that goes beyond the information in the premises (Khemlani & Johnson-Laird, 2011). Or else one could make a *modus tollens* deduction, which is a logical deduction that takes the following abstract form. *If A then B. Not B. Therefore, not A.* The inference is valid, i.e., the conclusion is true whenever the premises are true, but it is difficult for naïve reasoners. Thus, in the present study, the robotic confederate's responses concerned either an explanation or else a *modus tollens* deduction. Half of the participants interacted with a confident confederate and the other half interacted with an uncertain one. If the effect of uncertainty on credibility applies to any sort of response, then there should not be an interaction between the type of inference and the confederate's confidence. In contrast, if the effect is unique to explanatory reasoning, then there should be no difference between participants' evaluations of confident and uncertain deductions, but there should be a difference in their evaluations of explanations.

## Method

*Participants, design, and procedure.* 45 native English-speaking participants were recruited though the same participant pool as in Experiment 1. None of them had received training in formal logic. They solved ten reasoning problems by engaging in a web-based chat interaction with a pseudo-autonomous robotic confederate (see Figures 1 and 2), and they were taught to use the interface using the same procedure as in the previous study. Their task was to read each problem aloud to the confederate, listen to the confederate's response, and then judge whether they agreed, did not understand, or disagreed with the response. On half of the problems, the confederate would produce an explanation, and on the other half, it would produce a deduction (see Appendix). Twenty-three participants interacted with a confederate that produced confident responses and the remaining interacted with one that produced uncertain responses. After completing the last problem, participants filled out the same post-experimental questionnaire that was described for Experiment 1.

## Results and discussion

Figure 4 presents the percentage of agreement to deductions and explanations as a function of whether the response was delivered by a confident or an uncertain confederate. Participants agreed with deductions almost at ceiling (87%) and accepted them reliably more often than they accepted explanations (63%; Wilcoxon test, $z = 3.8$, $p < .0001$, Cliff's $\delta = .55$). Likewise, they accepted uncertain responses more often than confident responses (81% vs. 71%; Mann-Whitney test, $z = 2.47$, $p = .01$, Cliff's $\delta = .43$). However, the main effect of confidence was driven entirely

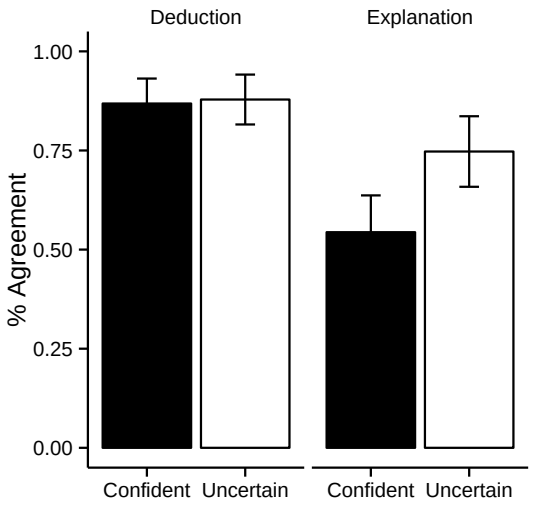



**Figure 4.** Agreement percentages for deductions and explanations as a function of whether they were delivered by a confident or an uncertain confederate. 95% confidence intervals shown.

by the difference between confident and uncertain explanations, and the data yielded a significant interaction between the type of inference and the confederate's confidence (Mann-Whitney test, $z = 1.95$, $p = .05$, Cliff's $\delta = .48$). The results suggest that the effect of uncertainty on credibility applies to explanations and not deductions. As in the previous study, agreement varied as a function of the contents of the problems (Friedman analysis of variance, $\chi^2 = 43.49$, $p < .0001$), and 58% of the participants reported that they believed the interaction was pre-recorded.

To assess whether the effect and the relevant interaction were both reliable across the different materials, we fitted the data to another generalized mixed-effects model. The model took into account two fixed effects, i.e., the confederate's confidence and the inference type, and the three pertinent random effects, i.e., the participant variance, the problem variance, and whether or not the participant believed that the interaction was pre-recorded. The model yielded a significant main effect of the type of inference ($b = -2.07$, $SE = .36$, $p < .0001$), however it yielded no main effect of confidence ($b = .05$, $SE = .42$, $p = .90$). Instead, it yielded a significant interaction between the type of inference and the confederate's confidence ($b = 1.07$, $SE = .54$, $p = .045$). As in Experiment 1, the analysis shows that the effect held in spite of any variance from the different materials or the perception that the interaction was staged.

# General Discussion

We used a novel experimental methodology to study how reasoners incorporate metacognitive information to judge one another's explanations. In two experiments, reasoners interacted with a robotic agent that appeared to deliver its responses in a confident or else an uncertain demeanor. One might expect that people should agree with confident explanations more often. Yet Experiment 1 showed that participants accepted explanations more often when they came from an uncertain confederate compared to a confident one. Experiment 2 tested whether the effect held

more generally for deductions, but it found instead that it was limited to explanations.

Why do reasoners accept explanations more often when they come from an uncertain source? The results are counterintuitive, particularly since confidence is correlated with informational accuracy. Indeed, at first blush, the results of our studies conflict with recent findings on hedging behavior and powerless language (Blankenship & Holtgraves, 2005; Durik, Britt, Reynolds, & Storey, 2008; Liu & Fox Tree, 2012). However, we hypothesize that one reason for a speaker to produce uncertain expressions, gestures, and verbal cues is to signal to a listener that the speaker is engaged in deeper analytic processing, and furthermore, that the speaker is considering alternative possibilities. This proposal accounts for why the effect is manifest for explanations but not *modus tollens* deductions: explanations require reasoners to think about multiple possibilities and to go beyond the information presented in the premises, whereas *modus tollens* deductions do not. If our hypothesis is true, then we should a similar effect of uncertainty on credibility for deductions that require reasoners to consider multiple possibilities compared to those that do not.

The present data reveal a robust credibility effect for human-robot interactions, and critics are justified in wondering whether the effect will still hold in dyadic human-human interactions (but cf. Moon & Nass, 1996, for evidence that people treat interactive computers as though they were human). Similar studies with human confederates are feasible, but the human-robot interaction paradigm we employed has several advantages to traditional studies with human confederates. First, robotic confederates can be programmed to yield very precise expressional and gestural cues that are consistent for all participants in the study, while even the best human confederates are susceptible to irregular behaviors. Second, robotic confederates can be programmed to implement complex experimental designs and counterbalancing schemes. For example, the software in Experiment 2 was written so that exactly half of the robot's responses were explanations. Despite these advantages, however, future studies should examine the credibility effect in, albeit less controlled, human studies. One promising methodological compromise is to run pseudo-dyadic interaction studies over the Internet (Summerville & Chartier, 2012).

The results we present have psychological implications, as well as implications for robotics researchers. A major goal for the interdisciplinary community of human-robot interaction research is to develop social robots that humans trust (Fong, Thorpe, & Baur, 2001; Goodrich & Schultz, 2007; Steinfield et al., 2006). The credibility effect we show implies that humans are likely to take into account metacognitive signals (and their robotic analogs) in assessing information from autonomous systems. Research on the modulatory effects of confidence on higher order reasoning is of multidisciplinary relevance, and can be applied to developing broader theories of confidence monitoring in humans as well as more natural and trustworthy autonomous robots.

## Acknowledgements

This research was funded by a National Research Council Research Associateship awarded to SK and ONR Grant #s N0001412WX30002 and N0001411WX20516 awarded to JGT. We are also grateful to Bill Adams, Len Breslow, Magda Bugajska, Tony Harrison, Laura Hiatt, Phil Johnson-Laird, Ed Lawson, Eric Martinson, Malcolm McCurry, Lilia Moshkina, Frank Tamborello, Alan Schulz, and the ARCH Lab at George Mason University for their helpful comments and their assistance in data collection.

## References

Albert, J.M., & Sponsler, G.C. (1989). Subjective probability calibration: A mathematical model. *Journal of Mathematical Psychology, 33,* 289-308.

Bates, D., Maechler, M., Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. Retrieved fromhttp://CRAN.R-project.org/package=lme4.

Blankenship, K. L., & Holtgraves, T. (2005). The role of different markers of linguistic powerlessness in persuasion. *Journal of Language and Social Psychology, 24,* 3-24.

Breazeal, C., Siegel, M, Berlin, M., Gray, J., Grupen, R., Deegan, P, Weber, J., Narendran, K., & McBean, J. (2008). Mobile, dexterous, social robots for mobile manipulation and human-robot interaction. In Proceedings of International Conference on Computer Graphics and Interactive Techniques. New York, NY: ACM.

Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review, 103,* 566-581.

Dunlosky, J., & Metcalfe, J., (2009). *Metacognition.* Thousand Oaks, CA: US Sage Publications, Inc.

Durik, A., Britt, M.A., Reynolds, R., & Storey, J. (2008). The effects of hedges in persuasive arguments: A nuanced analysis of language. *Journal of Language and Social Psychology, 30,* 341-349.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous overconfidence and conservatism in judgment: implications for research and practice. *Psychological Review, 101,* 519-27.

Ferrell, W.R., & McGoey, P.J. (1980). A model of calibration for subjective probabilities. *Organizational Behavior and Human Performance, 26,* 32-53.

Fong, T., Thorpe, C., & Baur, C. (2001). Collaboration, dialogue, and human-robot interaction. In *Proceedings of the 10th International Symposium of Robotics Research.*

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506-528.

Goodrich, M., & Schultz, A.C. (2007). Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.) *Syntax and semantics*, pp. 41–58. New York: Academic.

Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology, 24,* 411-435.

Hiatt, L.M., Harrison, A.M., Lawson, W.E., Martinson, E., & Trafton, J.G. (2011). Robotic secrets revealed, Episode 002 [Video file]. Retrieved from http://goo.gl/ru28e.

Johnson-Laird, P.N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review, 111.*

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57.*

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology, 57,* 225-254.

Khemlani, S., & Johnson-Laird, P.N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, 64, 2276-88.

Khemlani, S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica, 139*, 486-491.

Khemlani, S., Sussman, A., & Oppenheimer, D. (2011). *Harry Potter* and the sorcerer's scope: Scope biases in explanatory reasoning. *Memory & Cognition*, 39, 527-535.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review, 119*, 80-113.

Lagnado, D. (1994). *The psychology of explanation A Bayesian approach.* Masters Thesis. Schools of Psychology and Computer Science, University of Birmingham.

Lee, N.Y.L., & Johnson-Laird, P.N. (2006). Are there cross-cultural differences in reasoning? In D.S. Macnamara & J.G. Trafton (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 459-464). Austin, TX: Cognitive Science Society.

Lichtenstein, S., Fischhoff, B., Phillips, L.D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgment under uncertainty Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Liu, K., & Fox Tree, J. E. (2012). Hedges enhance memory but inhibit retelling. *Psychonomic Bulletin & Review, 19*, 892-898.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences, 10*, 464-470.

Lombrozo, T. (2007). Simplicity and probability in causal explanations. *Cognitive Psychology, 55*, 232-257.

McClelland, A.G., & Bolger, F. (1994). The calibration of subjective probabilities: theories and models 1980-1994. In G. Wright and P. Ayton (Eds.), *Subjective Probability*. Chichester: Wiley.

Moon, C., & Nass, C. (1996). How "real" are computer personalities? *Communication Research, 23*, 651-674.

Pfeifer, P.E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58*, 203-213.

R Core Team (2012). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., & Goodrich, M. (2006). Common metrics for human-robot interaction, in: Proceedings of the 1st ACM/IEEE International Conference on Human-Robot Interactions, HRI'06.

Summerville, A., & Chartier, C.R. (2012). Pseudo-dyadic "interaction" on Amazon's Mechanical Turk. *Behavioral Research Methods*, online publication.

Wallsten, T.S., & Gonzáles-Vallejo, C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review, 101*, 490-504.

Wilson, D., & Sperber, D. (2004). Relevance theory. In G. Ward and L. Horn (Eds.) *Handbook of pragmatics*, pp. 607–32. Oxford, UK: Blackwell Science.

**Appendix.** The problems used in Experiment 1 and Experiment 2, which consisted of a conditional generalization (column 1) and a categorical statement (column 2).

| **Premises (spoken by the participant to the confederate)** | | **Responses (spoken to the participant by the confederate)** | |
|---|---|---|---|
| **Conditional generalization** | **Categorical** | **Explanation (Experiments 1 and 2)** | **Deduction (Experiment 2)** |
| If a person is bitten by a viper then he will die | However, a man named Matthew did not die | Matthew received an antidote | Matthew was not bitten by a viper |
| If James does regular aerobic exercises then he strengthens his heart | But, James did not strengthen his heart | James had a congenital heart defect | James did not do regular aerobic exercises |
| If a car's engine is tuned in a special way then its fuel consumption goes down | However, one car's fuel consumption did not go down | The car had engine problems that increased consumption | The car's engine was not tuned in the special way |
| If the aperture on a camera is narrowed, then less light falls on the film | But in one instance, less light did not fall on the film | It was completely dark, so there was no light at all | The aperture on the camera was not narrowed |
| If a person pulls the trigger on a pistol, then the pistol fires | However, it turned out that the pistol did not fire | The safety had not been taken off the pistol | Nobody pulled the trigger |
| If a substance such as butter is heated then it melts | However, one piece of butter did not melt | The heat was too low to melt the butter | The piece of butter was not heated |
| If Chemical A and Chemical B come into contact with one another then there will be an explosion | But there was no explosion | There was not enough of either of the substances | The two substances did not come into contact with one another |
| If a person receives a heavy blow to the head then that person forgets some preceding events | However, Pat did not forget any preceding events | Pat was wearing a helmet at the time | Pat did not receive a heavy blow to the head |
| If people make too much noise at a party then the neighbors complain | But the neighbors did not complain | The neighbors were away on summer vacation | People did not make too much noise at the party |
| If the banks cut interest rates then the GDP increases | But the GDP did not increase | Cutting rates is not enough in an economic decline | The banks did not cut interest rates |